

Small or medium-scale focused research project (STREP)



**ICT Call 8**

FP7-ICT-2011-8

**Cooperative Self-Organizing System for low Carbon Mobility at  
low Penetration Rates**

**COLOMBO**

**COLOMBO: Deliverable 5.3**

**Traffic Light Algorithm Evaluation System**

Document Information	
Title	Deliverable 5.3 - Traffic Light Algorithm Evaluation System
Dissemination Level	PU (Public)
Version	1.0
Date	10.11.2014
Status	Final version

Authors	Daniel Krajewicz (DLR), Robbin Blokpoel (IMTECH), Wolfgang Niebel (DLR), Ting Lu (DLR), Cornelia Hebenstreit (TUG), Martin Rexeis (TUG), Laura Bieker (DLR)
---------	---

## Table of contents

1	Introduction.....	5
1.1	Document Objectives .....	5
1.2	Document Structure.....	6
2	State of the Art.....	7
2.1	Official Procedures for TLC Evaluation .....	7
2.2	Existing TLC Evaluation Software .....	12
2.3	Criteria and Performance Indicators.....	13
2.4	Discussion .....	14
3	Relevant Characteristics of Real-World Traffic .....	15
3.1	Traffic demand load curves.....	15
3.2	Short-Term Traffic Flow Variability.....	22
3.3	Turning Ratios .....	26
3.4	Vehicle Fleet Compositions .....	28
4	New Methodology for Evaluating Traffic Light Systems .....	29
4.1	Performance Indicator Selection .....	29
4.2	Development of a Single Measure .....	31
4.3	Implementation.....	34
5	TLS Algorithm Evaluation System .....	35
5.1	Requirements.....	35
5.2	Realisation .....	36
5.3	Scenario Sets .....	39
5.4	Traffic Participants .....	41
5.5	PI Computation.....	44
5.6	Implementation and Deployment .....	46
5.7	First Insights .....	48
6	Summary.....	50
	Appendix A – References .....	51
	Appendix B – Load Curves Clustering Dendrograms .....	54
	Appendix C – Scenario Sets.....	55



# 1 Introduction

The COLOMBO project develops a set of modern cooperative traffic surveillance and traffic control applications that target at different transport related objectives such as increasing mobility, resource efficiency, and environmental friendliness. The COLOMBO project develops applications only relying on simulations. If these applications prove to bring benefits to the traffic system within the simulation, further steps will get necessary to realize a real-world implementation of these systems.

Thereby, a valid evaluation methodology is one of the key techniques used and addressed within COLOMBO. The evaluation of traffic surveillance algorithms using simulations is relatively straight-forward as the algorithms can be compared with the ground-truth, which is available within the simulation. However, for advanced traffic light control this is more complex. The perfect control strategy is not known, so a comparison with a ground-truth is not possible. Additionally, a large number of different stakeholders may influence the decisions while designing a traffic light control and a large number of possibilities to benchmark a traffic light exist.

COLOMBO's Work Package (WP) 1 revealed the need for a common scientific methodology to appraise and benchmark traffic light control (TLC) algorithms under different circumstances and in different projects and countries (cf. [COLOMBO D1.1, 2014]). No unified approach for measuring the benefits of a developed TLC could be found. In contrary, the examined reports use a large variety of performance indicators and individual scenarios with their differing road networks, demand volumes, and TLC parameters (number of phases, cycle time). Such heterogeneous evaluations make the results hardly comparable. To address this issue, generally applicable scenarios were modelled for being released to the long-lasting public usage. They aim on enabling an in-depth understanding of the developed traffic light algorithms' behaviour. The need for open accessible sample scenarios to test simulation software's conformity to official evaluation procedures was addressed, i.a., by the US-American Committee on Highway Capacity and Quality of Service (HCQS) [Kittleson and Roess, 2001].

Besides proper scenarios, meaningful performance indicators that describe the algorithm's benefits have to be used for the evaluation. The increasing number of available performance indicators that address different sub-topics, such as traffic efficiency, environmental impacts of traffic, or road user perception, may reduce the expressiveness of obtained simulation results as a reader may need to choose the measure she/he is interested in. Therefore, a single performance indicator (PI) that considers the named sub-topics was developed.

This deliverable describes the resulting artefacts, a test execution system that helps in evaluating new traffic light control algorithms and that is based on a large variety of scenario (sets), and a performance indicator that incorporates all relevant aspects of a traffic light's performance. Both are realised as individual software packages but may be used in conjunction as the test execution system can execute the single PI computation.

## 1.1 Document Objectives

The objective of this document is to present how selected parameters of the scenarios were derived by real-world investigations. It also outlines the choice of performance indicators and their further transformation within the evaluation process for traffic lights algorithms. The description of the general work flow and its specific implementation enables potential users of the software system to apply, configure, and execute the software.

The intended audience are practitioners like TLC designers who wish to comprehensively evaluate their work; traffic planners and managers who wish to clearly structure and set out the evaluation framework of their transport development plans; and programmers of microscopic traffic simulation software who want to include the scenarios as input, and dock their outputs onto the evaluation part.

## **1.2 Document Structure**

At first, a summary on how traffic lights are benchmarked nowadays is given in Chapter 2. Then, some characteristics of the behaviour of traffic participants at traffic lights are presented in Chapter 3. Chapter 4 introduces a newly developed methodology for evaluating traffic light systems. Chapter 5 then presents the system that helps in evaluating traffic light algorithms. Chapter 6 summarises the work.

## 2 State of the Art

The methodology of impact assessment and evaluation was already described in [COLOMBO D1.1, 2014], section 2.4. The therein briefly mentioned evaluation procedures will be more extensively presented with their elements, advantages, and shortcomings for usage in COLOMBO in section 2.1 of this chapter. Section 2.2 gives a brief overview about existing evaluation software, which shall allow to estimate how desired and successful an independent TLS algorithm evaluation application could be in practice. Section 2.3 contains new PIs which are to be included into the new evaluation procedure of chapter 4. This chapter ends with a summary.

### 2.1 Official Procedures for TLC Evaluation

#### *Austrian Guideline RVS 05.04.35*

In Austria for TLC evaluation the guideline RVS 05.04.35 [FSV, 2013] with its procedure EVA (Evaluierung von Verkehrslichtsignalanlagen/Evaluation of TLA) is used for both newly erected light signals and significantly modified ones. No longer than two years after the installation, the first evaluation must be finished. Afterwards, an evaluation has to be performed biennial. The EVA procedure follows the method of a multi-criteria-analysis (MCA).

The evaluation targets at the objectives of traffic safety and security as well as at traffic performance, which are the reasons to install a TLC according to §36/1 StVO [BMVIT, 2014]<sup>1</sup>. Beside these objectives the evaluation implicitly addresses user satisfaction, which is understood as a separate goal (cf. [TP 2007], [FESTA 2008]). These global goals are broken down into eight criteria (“check lists”) and further operationalized by sets of indicators. The measurements or observations of these indicators are transformed into values of an ordinal scale with the range between 1 (desirable) and 3 (undesired), whereat sometimes only dichotomous pairs are possible, e.g., the criterion is “present” (ordinal 1) vs. “absent” (ordinal 3). The importance of each indicator and hence its weight within its set is equal; thereby the arithmetic average is determined. To calculate the resulting single decision value the criteria are weighted. Objectives, criteria, their suggested weights, and indicators are given in Table 2.1. It is noteworthy that some criteria are not independent from each other although evaluation theory demands this, e.g., list 2 (Perceptibility/Comprehensibility) hints at conflict situations of list 1 (Accidents). List 6 does not contain any explicit indicator for environment friendliness, but argues that this is correlated to traffic performance and its indicators of list 3.

**Table 2.1: Goals, criteria, weights, and indicators of the Austrian Evaluation Procedure.**

Objective	Criteria	Weights	Indicators
Safety	Accident Occurrence (List 1)	0.20	Accidents with injuries in last 2 years; Accident black spot (y/n); Similarity of accidents; Occurrence of conflict situations <sup>2</sup>
	Perceptibility / Comprehensibility (List 2)	0.10	Completeness and unambiguity of signal heads, markings, and signage; Lighting; Conflict situations <sup>2</sup>

<sup>1</sup> „Die Behörde hat zur Wahrung der Sicherheit, Leichtigkeit und Flüssigkeit des Verkehrs auf Straßen mit öffentlichem Verkehr unter Bedachtnahme auf die Verkehrserfordernisse zu bestimmen, ob und an welcher Stelle der Verkehr durch Armzeichen oder durch Lichtzeichen zu regeln ist“ („For the sake of safety, ease, and fluidity of traffic on public roads the authority has to decide under consideration of the traffic demands, if and where to regulate traffic manually or by traffic lights.“)

<sup>2</sup> Conflict situations are addressed in lists 1, 2, and 5.

Security	Equipment Reliability / Availability (List 7)	0.15	Number of failures per year
	Equipment Condition (List 8)	0.10	State of 6 different TLS subcomponents (repair needed vs. ok)
Traffic Performance and Mobility	Motor Traffic (Lists 3 & 6)	0.10 & 0.10	Average waiting time; number of stops; Coordination with subsequent junctions; External influences; Demand-capacity-ratio; Queuing space
	Public Transport (list 4)	0.10	Average waiting time; External influences
	Pedestrians & Cyclists (List 5)	0.15	Average and maximum waiting time; Conflict situations <sup>2</sup>
User Satisfaction and Acceptance	Ease and Comfort of Use (partly lists 4 and 5)	Part of other weights (0.10 & 0.15)	Access to public transport stop; Adaption to special needs of people; Waiting space; (Perception of) traffic performance indicators

By utilizing such a consistent evaluation procedure, the results are comparable at least within one administration unit. To gather sufficient input data documentation as component of quality management is needed.

It becomes clear that some of these criteria and indicators cannot be used in a simulation-based evaluation, as no model to represent them is implemented. This counts for both of the security criteria, the safety criterion Perceptibility / Comprehensibility, and most of the user satisfaction indicators. Microscopic modelling of the safety-related criteria accidents and conflicts is still at an early stage and disputed, namely the Surrogate Safety Assessment Method (SSAM) [FHWA, 2008]. For normative evaluation a radar chart like in Figure 2.1 is also used.

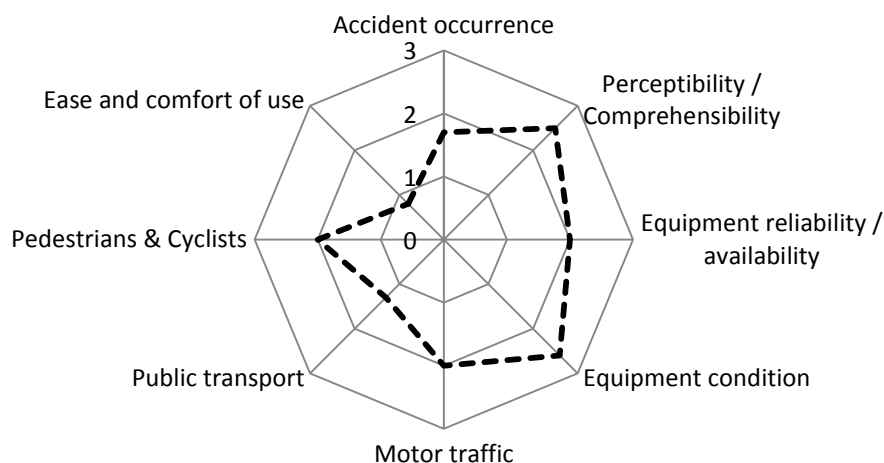


Figure 2.1: Exemplary 8-criteria radar chart according to RVS.

### Germany Manual HBS 2009

The “Handbuch für die Bemessung von Straßenverkehrsanlagen” (HBS) [FGSV, 2009] is a comprehensive manual for the design of road infrastructure and also includes assessment and evaluation procedures for most types of roads, i.e., highways, rural roads, and junctions of urban



roads. In German-speaking countries, it is one of the most important standard works since its release in the year 2001. The great advantages of the therein applied evaluation method are its standardization and simplicity due to a uniform procedure structure. That is, transforming a single representative indicator of the traffic performance into one of the six levels of service (LoS). The indicator values can be determined by real measuring, by applying the HBS calculation models, or via user-chosen methods like microscopic simulation. The calculation is supported by a form as tool.

The LoS ranges between A (totally uninfluenced driving or walking on the free section, virtually no or very short waiting times) to F (critical). For each type and element of a road and if applicable for each transport mode (individual motor vehicles, on-street public transport, cyclists, and pedestrians) different indicators and different value ranges are applied. There is no aggregated LoS neither of all involved transport modes nor of successive road elements like a signalised arterial. This leads to the deficit of a piecewise evaluation without a clear aggregated overall result.

The performance indicator for all four regarded transport modes at signalized intersections according to HBS Chapter 6 is the average waiting time. It is stated separately for each lane/signal group and often uses the peak hour as its time denominator. In addition, for coordinated TLS corridors (“green wave”) the indicator for motor vehicles is the percentage of unstopped passing. The average waiting time of motor vehicles is composed of the basic latency due to red time, and the time loss due to congestion feedback, which is dependent on the junction saturation degree. Table 2.2 shows the thresholds of both indicators and their according Level of Service.

**Table 2.2: Indicator threshold values for Level of Service (HBS).**

Level of Service	Average waiting time $t_w$ [s]				Percentage of passing without stop [%]
	On-street public transport	Bicycles	Pedestrians	Motor vehicles (uncoordinated)	Motor vehicles (coordinated)
A	$\leq 5$	$\leq 15$	$\leq 15$	$\leq 20$	$\geq 95$
B	$\leq 15$	$\leq 25$	$\leq 20$	$\leq 35$	$\geq 85$
C	$\leq 25$	$\leq 35$	$\leq 25$	$\leq 50$	$\geq 75$
D	$\leq 40$	$\leq 45$	$\leq 30$	$\leq 70$	$\geq 65$
E	$\leq 60$	$\leq 60$	$\leq 35$	$\leq 100$	$\geq 50$
F	$> 60$	$> 60$	$> 35$	$> 100$	$< 50$

A new release of the HBS is under work, which might yield changed threshold values as well as different performance indicators. The general standard to keep all procedures rather simple to be understood, interpretable, and – at least theoretically – calculable for humans, will be retained.

### *German Guideline RiLSA 2010*

The “Richtlinie für Lichtsignalanlagen” (“RiLSA”, [FGSV, 2010]) as the German guideline for TLS design lists as possible goals for a TLS erection the improvement of traffic safety, traffic performance, environment friendliness (emissions, land use), and efficiency (fuel consumption). It points out that the involved stake- and shareholders might have conflicting objectives and expectations, which need to be balanced out according to their importance. In chapter 8 “quality management” the determined key performance indicators of the traffic performance are “waiting time” and “number of stops” from which further indicators like journey times, fuel consumption, noise emissions, and pollutant immissions can be derived. Traffic safety indicators are the number

and severity of traffic accidents, based on the accident-cost-rate and the accident density. A method to handle these indicators is not given in the RiLSA. It rather refers to the HBS and accident analysis code of practice.

### *German Recommendations EWS 1997/Guidelines RWS 2015*

The comprehensive German Cost-Benefit-Analysis procedure EWS [FGSV, 1997] is suitable for small-scaled road infrastructure construction appraisals. It comprises eight criteria of benefit, defined as the difference between the criterion's values of the base scenario and a comparison scenario. A project is worth to be realized when the benefits are greater than the costs. Generally speaking some of the criteria and their cost unit rates - to transform (monetise) the criterion's original dimension into a monetary value - have been proved to be suitable also for "non-hardware" projects such as (cooperative) intelligent transportation systems (C-ITS) (cf. [Niegel, 2013]). As successor the guidelines RWS [FGSV, 2015] are in work.

### *US-American Highway Capacity Manual HCM2010*

The evolution of the Highway Capacity Manual since 1965 with its concept of a "Level of Service" (LoS) grade is well explained in [Smart et al., 2014]. The newest HCM edition [TRB, 2010] features an additional multimodal framework to mirror the interrelationships between different modes of transport, but does not support a single multimodal LoS (MMLOS). Multiple performance indicators, depending on the transport mode, are mixed with spatial design parameters to calculate the respective LoS. This makes procedures and results harder to interpret and puts additional requirements onto data input and output of traffic simulations.

At signalized intersections the motorized vehicles' LoS is a simple grading function of the average vehicle control delay. It may be calculated per junction, per approach, or per lane group. Pedestrians and cyclists get scores to which the PI "waiting time" contributes only partly, next to driven vehicle speeds on the road, traffic volumes, the geometric design, and even the percentage of occupied on-street parking space, which could be interpreted as how comfortable users feel and thus accept the facility. The threshold values are listed in Table 2.3.

**Table 2.3: Indicator threshold values for Level of Service (HCM).**

LoS	Average waiting time $t_w$ [s] of motor vehicles	Score [-] of pedestrians and bicycles
A	$\leq 10$	$\leq 2.00$
B	$\leq 20$	$\leq 2.75$
C	$\leq 35$	$\leq 3.50$
D	$\leq 55$	$\leq 4.25$
E	$\leq 80$	$\leq 5.00$
F	$> 80$	$> 5.00$

### *TRANSYT Performance Index PI*

The bi-criteria Performance Index  $PI^3$  was implemented in the late 1960's into the TLC optimisation software TRANSYT by the Transport Research Laboratory (TRL). It is also used, i.a., by [Wietholt, 2009] and synthesises the waiting time  $w$  and the number of stops  $h$  of all traffic modes  $z$  and all access sections  $i$  of an intersection (Eq. 2-1). The weight  $Gh$  is assumed to be 60, since the emissions of a start-up after a stop equal 60 seconds idling. In difference to the HBS not vehicles but passengers  $P$  are used, thus incorporating the occupancy rate. Normally the peak hour

<sup>3</sup> Not to be confused with the PI for "Performance Indicator".

is used as time denominator, while the areal denominator can span from a single junction to a whole network.

$$PI = \frac{\left( \sum_i \sum_z w_{i,z} \times P_{i,z} + G_h \sum_i \sum_z h_{i,z} \times P_{i,z} \right)}{\sum_i \sum_z P_{i,z}} \quad (2-1)$$

### *Practical Approach in the Netherlands*

In the Netherlands no clear standard like the HCM or HBS is available for evaluation of traffic light controllers. In general every road authority has specific preferences and policies for each part of the network. Therefore, the evaluation criteria are customized for each project. This customization is mostly on how much weight is put on certain criteria while the general framework remains the same. The measurements that are usually acquired on a per signal group basis are the following:

- Delay time, the difference between free flow travel time and the actual travel time
- Number of stops, whenever a vehicle drives slower than 5 km/h it is considered a stop
- Queue length, only used when a maximum queue length is required due to potential road blockage upstream.
- Cycle time, used to ensure the first vehicle behind the stop line never needs to wait longer than that cycle time before it gets green.
- Demand waiting time, a more modern version of the cycle time criterion. This measures the time between the first vehicle stop at the stop line and the time the signal turns green. This gives extra flexibility for the controller to wait longer to give green when a vehicle arrives long after the light turned red.

In general, the total average delay time over all signal groups and intersections is the leading indicator, while other indicators are used more like constraints. For instance, a maximum cycle time of 100 seconds can be demanded in order to prevent road users from violating the red light because the waiting takes too long. The same holds for queue length; any strategy that would exceed a maximum queue length specified would not be accepted. The number of stops is used as an indicator for travel comfort and pollutant emissions. Clear guidelines to evaluate this are, however, not provided.

Often the performance of traffic controllers is also evaluated on a route level. In this case the amount of stops are more important, for instance a constraint can be that on a corridor of 4 intersections vehicles should not stop more than once on average. These routes consider either public transport or vehicles and get special attention in the overall evaluation regarding their delay. For public transport this delay may even be a constraint.

### *Other Countries*

In Switzerland the basic norm SN640 017a [VSS, 1999] and the particular norm SN640 023a for signalised junctions [VSS, 2008] contain the LoS concept similar to HCM and HBS.

A survey posted in a couple of interest groups on the professionals' network LinkedIn yielded in five answers from different countries, but rather assuring the topic of TLC evaluation was considered as important than giving substantial insight into practitioners' approaches.

## Overview

The following Table 2.4 gives an overview about most of the different presented traffic evaluation procedures and their contained criteria.

**Table 2.4: Overview on traffic evaluation procedures and contained criteria.**

Objectives	Criteria	Procedure				
		RVS	HBS	HCM	PI	EWS
<b>Performance and Mobility</b>	Travel Time including delay and stop / Waiting Time	x	x	x	x	x
	Number or Percentage of Stops	x	x		x	
<b>Environmental Impact</b>	Pollutant Emissions (NO <sub>x</sub> , CO, HC, PA)					x
	Climate Gas CO <sub>2</sub>					x
	Noise Emissions					x
<b>Resource Efficiency</b>	Fuel Consumption					x
	Building / Acquisition Costs					x
	Operating + Maintenance Costs					x
	Occupancy Rate				x	x
<b>Safety</b>	Perceptibility / Comprehensibility; Accident Occurrence	x				
<b>Security</b>	Equipment Condition and Reliability	x				
<b>User Satisfaction</b>	Ease and Comfort of Use	x		x		

## 2.2 Existing TLC Evaluation Software

Traffic light control (TLC) evaluation software often comes as a component of TLC design software or of microscopic traffic simulations. While the former apply analytical calculation methods to generate TLC programs and the resulting PI values deterministically, the latter form a model of a highly dynamic system with many interactions to derive the PI values of a given TLC algorithm in a more stochastic way. The results acquired from the simulators can then be either directly processed by the software's evaluation module or stored in external files. These files are either given to external evaluation tools, or manually opened and read in a text editor. To obtain the results when no external tool is used, the user has to specify which data to acquire, first. This is done through addition of detectors, travel time sections and queue counters to the simulation network layout. Post processing of vehicle log files to get more or more accurate PIs, e.g., pollutant emissions, is possible as well.

Commercial TLC design products like Sitraffic Office (Siemens), Ampel (BPS GmbH), and LISA+ (Schlothauer & Wauer) as well as the simulation software VISSIM (PTV) refer to the German manual HBS. The evaluation suite MAT.CrossCheck (MAT.Traffic) is designed to interact with VISSIM and comprises the procedures of the HBS, too. VS-Plus (Verkehrs-Systeme AG) can be evaluated only by coupling it to a VISSIM simulation. The simulation software AIMSUN (TSS) and the design software TRANSYT (TRL) include algorithms to compute the LoS according to the US-American HCM. For the Dutch evaluation, which is different for each road operator, the simulation result files are interpreted manually with a text editor or a spreadsheet program like Excel to apply the weights specific to the network.

## 2.3 Criteria and Performance Indicators

Beside the criteria and PIs already listed in [COLOMBO D1.1, 2014] chapter 3, some new possibilities to address the additional global objective “User Experience” will be presented here.

### *Waiting Time Acceptance (Pedestrians and Cyclists)*

A literature study about waiting time perception for pedestrians was carried out in [Martin, 2006]. It found several sources where pedestrians are reported to get impatient at around 25-30 seconds waiting time. Additionally after 40 seconds the risk of red light violation was reported to increase substantially.

For cyclists [Yang, 2012] found that 32 % can be identified as risk takers and would cross whenever possible as their waiting endurance is less than 3 seconds. In the same study 53 % would wait at most 30 seconds and the last 15 % can be considered risk-averse. Two other sources [PRESTO, 2009] and [Fietsberaad, 2004] recommend a maximum waiting time of 90 seconds.

### *Waiting Time Perception (Drivers)*

Every road user experiences his waiting time differently than the physical waiting time. But for evaluating the waiting time and for measuring the acceptance of a traffic light it is also important to consider the perceived waiting time. Usually the perception of traffic lights is not included in the process of developing a traffic light controller. For this reason [Bijl et al., 2011] analysed the topic of perception of waiting time at signalized intersections. In the study the perceived waiting time  $t_{waiting}^{perceived}$  is a function of the actual waiting time, depicted in equation 2-2 and parameterised according to Table 2.5. Additional factors that influence the drivers experienced waiting time are the number of stops and the presence of a red wave. The number of stops ( $n_{stop}$ ) is defined as the number of times a car has stopped in the same queue. A red wave (RW=1) or no red wave (RW=0) depends if the car has to stop at two or more consecutive intersections or not. A red wave at the first intersection is naturally not possible.

$$t_{waiting}^{perceived} = \beta_0 + \beta_1 \cdot RW + (\beta_2 + \beta_3 \cdot n_{stop} + \beta_4 \cdot RW) \cdot t_{waiting} + \beta_5 \cdot t_{waiting}^2 \quad (2-2)$$

**Table 2.5: Static parameters for the perceived waiting time.**

Parameter	Value	p-Value
$\beta_0$	13,859	-
$\beta_1$	17,254	0.11
$\beta_2$	0,661	0.00
$\beta_3$	- 0,233	0.00
$\beta_4$	- 0,432	0.03
$\beta_5$	0,006	0.00

Interestingly the value for  $\beta_0$  is applied even when the driver could pass without stopping. This is because it turned out that drivers prefer to wait a few times shortly over waiting once for a longer time. If there is a next intersection, however, then  $\beta_1$  offsets this preference again because of the red wave. Therefore, it can be concluded that on a long stretch of roads with multiple intersections, a driver prefers to wait shortly every other intersection rather than a long wait at the beginning followed by an extended green wave.

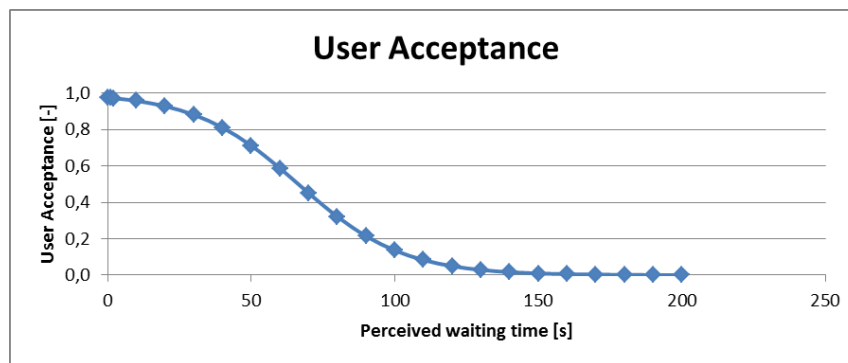
### Waiting Time Acceptance by Drivers

Furthermore, the study focused on the relationship between perceived waiting time and the user acceptance. The user acceptance (UA) of signalized intersections is considered as a function of perceived waiting time expressed in equation 2-3 with parameters in Table 2.6. The UA can adopt values from 0 to approximately 0.95 in practice as shown in Figure 2.2. As this PI is expressing the average presumption of how the road user is going to accept the waiting time, it remains a mystery of the study authors why no values between 0.95 and 1.00 can be achieved.

$$UA = \frac{1}{1 + e^{\beta_0 + \beta_1 \cdot t_{\text{waiting}}^{\text{perceived}}}} \quad (2-3)$$

**Table 2.6: Static parameters for the user acceptance.**

Parameter	Value
$\beta_0$	- 3,650
$\beta_1$	0,055



**Figure 2.2: The PI “User Acceptance” as a function of the perceived waiting time.**

## 2.4 Discussion

Neither common procedures nor a comprehensive selection of indicators are established on a supranational level. Even on a national level only a few countries have defined these, with deviations being allowed.

The grading LoS method to transform PI results from a continuous ratio or interval scale into a step-wise defined ordinal scale with its inherent loss of information and even leading to probable wrong conclusions was criticized by [Kittleson and Roess, 2001] and exemplary shown, e.g., in [Hunter, 2010]. Another criticism is the independence of thresholds from (locally) differing user perceptions (city centre vs. rural). These concerns were broadened by [Smart et al., 2014] onto the aggregation of PIs in multimodal and multi-criteria methods.

There is also a raising request to differentially regard the user and put his satisfaction onto the list of objectives. Beside the presented approaches in section 2.3, the Australian “Traffic Frustration Index” [Akcelik, 2000] can be named.



### 3 Relevant Characteristics of Real-World Traffic

As comprehensively described in [COLOMBO D1.1, 2014]<sup>4</sup>, scenarios are constituted by objects and their parameters of spatial, temporal, regulatory, behavioural, and technological nature. Some of the most relevant parameters and their particular occurrence in the real world are further investigated in the following. This comprises the (spatio-)temporal representation of traffic demand as long-term load curves, short-term traffic variability, and turning ratios at junctions; the technologically reasoned fleet distribution of different vehicle types according to their combustion engines.

The gained insights trigger the systematic build of synthetic scenario sets in Chapter 5, which resemble the real world by applying only those parameter values and their combinations that were found to be significantly differing.

#### 3.1 Traffic demand load curves

##### *Motivation for load curves and their typification*

Road infrastructure and traffic light controllers (TLC) are often designed to cope with the traffic demand  $q$  of the  $n$ -th busiest (peak) hour of a year, with  $n=30$  in, e.g., Germany [FGSV, 2009] and the USA. That means, the 29 even busier hours will still be oversaturated, while a lot of hours at the other end of the annual scale (with 7.680 hours) have a much lower demand. Whether the TLC policy/algorithm under investigation is suitable for all year round strongly depends on how it fits these other demand (and saturation) levels. To investigate this question traffic load curves need to be applied.

Since full data availability is rarely given for the infrastructure part under investigation, extrapolation and temporal scaling upon few (manually) counted hours is necessary. This is supported by different space- and time-dependant scaling factors, which can also be presented as curves: either monotone cumulative load duration curves in which the  $n$ -th busiest hour simply ranks on the  $n$ -th place of the x-axis, or as load curves with inflection points. The latter one is described in the following.

##### *What are load curves?*

Traffic load curves represent the timeline of the traffic demand on a certain place. Commonly used is the absolute number of traffic participants per time interval, or their relative share in [%] throughout the considered period of the curve. Typical load curve periods are one day (24 h), one week (7 d; 168 h), or a whole year (365 d; 8.760 h). The respective time intervals for the daily and weekly curves' resolution is often one hour, for daily curves even going down to 5 minutes. Annual curves have the share per day or a correction factor which is applied to the Average Daily Traffic (ADT).

Traffic flow changes of the aggregated hourly period can be explained by **systematic** changes of the travellers' traffic demand, which is caused by the seasonal weather, the type of day (working day, holiday, weekend), the time of day, the resulting activities such as commuting, shopping, leisure, school attendance, praying (in regions where religion plays a big role), and the mode and route choice. Short-term flow variability in intervals less than an hour is more stochastic due to overlapping individual **random** behaviour [Lämmer, 2014] and described in section 3.2.

Load curves can be produced for different traffic participant types such as people or vehicles, private cars, or heavy goods vehicles (HGV). Load curves might represent the bi-directional road

<sup>4</sup> section 2.2 "Scenario Classification"

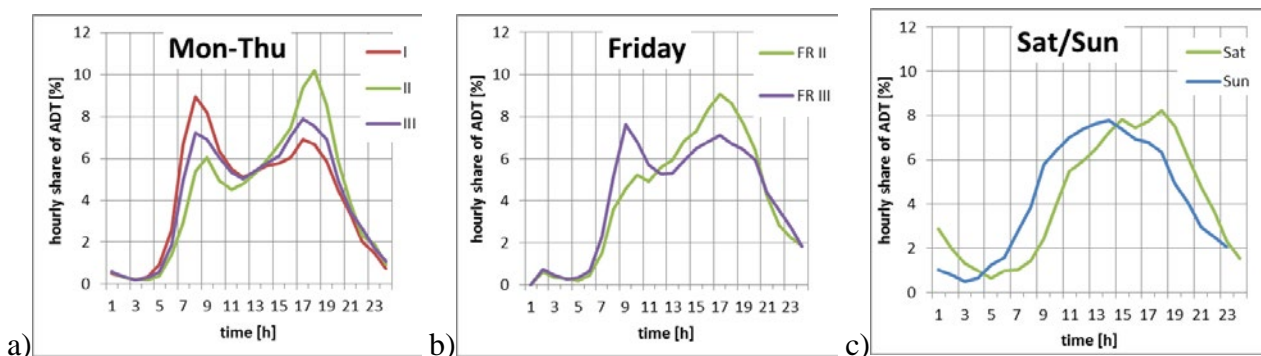
section or just one direction. Real curves arise wherever a long-term counting takes place in a particular space. Clustering over numerous normalized real curves leads to different “synthetic” curve *types*, which distinguish from each other significantly and cannot be exactly found in the real world. Road curve types always state the relative share in [%].

### *Load Curve Types (Motorized Vehicles)*

The following section describes uni-directional load curve types found in German, Suisse, and US-American literature and guidelines. It is shown further down in this section on the example of Bologna in Italy, that they generally can be applied at least also in other European countries, but daily curves might have a shift on the time-axis due to different day-plans. The focus of the COLOMBO project on urban roads is mirrored in this section by limiting the remarks mainly on this type of infrastructure.

The most recent reference which contains load curve types is the drafted German Guideline “Richtlinien für Wirtschaftlichkeitsuntersuchungen an Straßen” (RWS) [FGSV, 2015], subject to official publication. While it distinguishes between passenger cars/LGV and HGV for interurban roads, load curve types for urban roads are identic for all vehicle classes. There is only one annual type with 365 factors for the ADT ranging between 0.4342 (1<sup>st</sup> January) and 1.3313 (work day in April). The thereby derived ADT volumes are further split into hourly traffic volumes by 3 load curve types valid from Monday till Thursday (Figure 3.1 a), two types valid only on Fridays (Figure 3.1 b), and one type each for Saturdays and Sundays (Figure 3.1 c). With regard to the stochastic approach further down it is noteworthy that this procedure leads to identic variation coefficients for each hour within the same category of day.

Type I Mon-Thu has a strong morning peak period (Figure 3.1). Type II Mon-Thu has a strong afternoon peak period; and Type III Mon-Thu has two almost equal peaks. Friday Type II has also a strong afternoon peak period but with a lower peak hour volume, and a two hours earlier rise than Mon-Thu. Friday Type III again has two peaks. Saturday and Sunday have a plateau spreading 8 to 10 hours rather than a peak, with Sunday shifted 2 hours earlier.



**Figure 3.1: Daily Load curve types for (a) Monday to Thursday, (b) Fridays, and (c) weekends [FGSV, 2015].**

The German “Bundesverkehrswegeplan” [BMVBW, 2003] does not explicitly distinguish between urban and interurban roads, but its main application is for large-scaled interurban road projects. ADT changes during a year are covered by two downsizing and two upsizing factors for the work day ADT, the holiday ADT, and the Sunday/holiday ADT respectively. It stages 18 work day types (A-R) for passenger vehicles, 3 for duty vehicles, and 1 Sunday/holiday type. The 18 work day types can be clustered in such a way to resemble the 3 types of the RWS Mon-Thu, but each with 6 representations of different magnitudes. For example the Type I with its strong morning peak period has peak hour volumes between 7.4 % and 11.0 % (Figure 3.2 a). The duty vehicle types show only one peak at different magnitudes, which is around noon and holds for 6.7% to 8.9 % (Figure 3.2 b).



The Sunday load curve type is depicted in Figure 3.2 c). It has only one peak in the afternoon and is thereby similarly shaped to the RWS Sunday type in Figure 3.1.

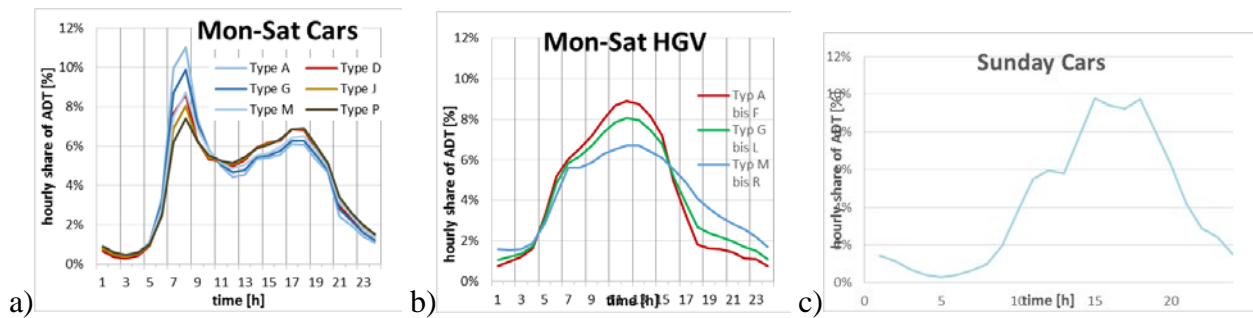


Figure 3.2: Daily load curve types for (a) passenger cars with morning peak, (b) HGV Mon-Sat, and (c) Sunday load curves [BMVBW, 2003].

An earlier investigation throughout the German road network [Pinkowsky, 2006] clustered annual, weekly, and daily load curves of passenger cars. The derived 6+1 work day types, shown in Figure 3.3, contain 2 representatives of each RWS type: Type I with a morning peak (A & B), Type II with an afternoon peak (E & F), and Type III with two smooth peaks (C & D). An additional morning peak type “G” occurs only on Mondays. Table 3.1 provides the few typifications with regard to road classes which could statistically viable be done.

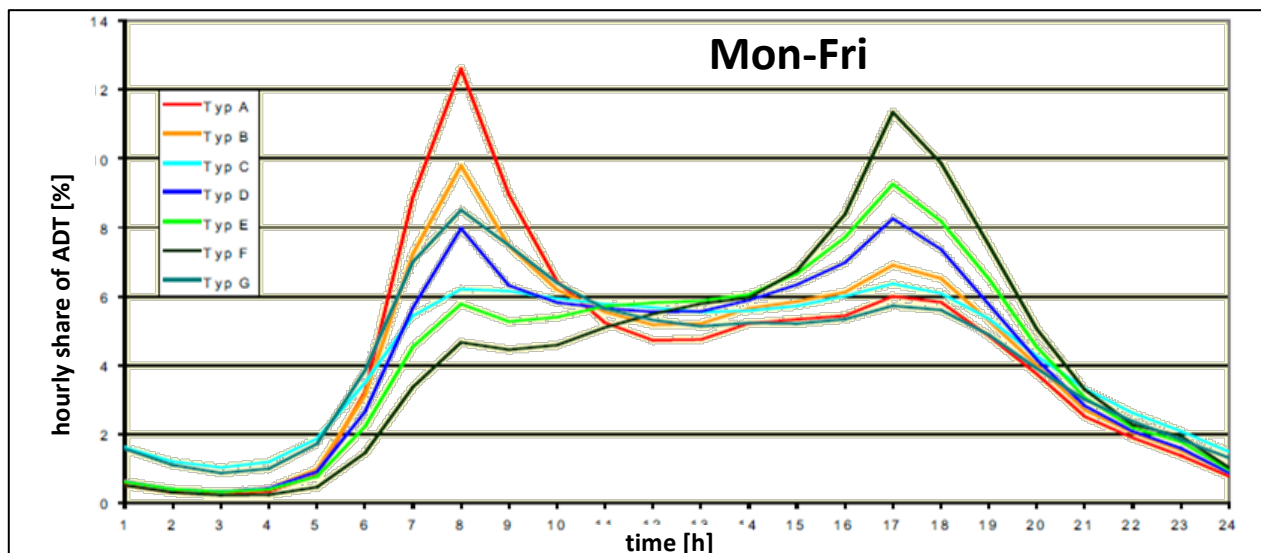


Figure 3.3: Work day load curves [Pinkowsky, 2006].

Table 3.1: Occurrence of load curve types [Pinkowsky, 2006].

Load Curve Type	Remark
A & B	Corresponds with types E & F; mainly commuter traffic
C	Mainly at motorways
D	Mainly on subordinate road network
E	Corresponds with types A & B; mainly commuter traffic
F	Corresponds with types A & B; mainly on subordinate road network; mainly commuter traffic
G	Only on Mondays, weekend commuters

The Suisse norm 640 005b [VSS, 2010] bases on the work of [Bernard and Axhausen, 2010]. Beside 5 annual curve types it contains 7 weekly curve types (7 d\*24 h) and thereby shows that

work day types strongly correlate with Saturday and Sunday curve types (Figure 3.4). Urban roads are hardly represented; “Gruppe 2”, “Gruppe 5”, and “Gruppe 6” are offside motorways, but might also represent rural areas.

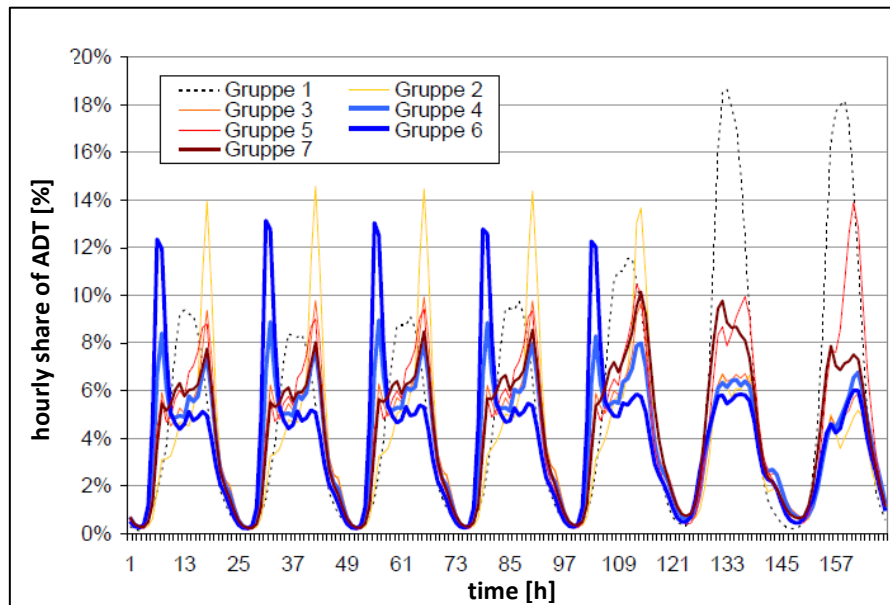


Figure 3.4: Weekly load curve types [Bernard and Axhausen, 2010].

Similar findings for the USA are reported in the HCM chapter 3 [TRB, 2010].

### *Load Curve Types (Bicycles and Pedestrians)*

A recent German research project [Schiller et al., 2011] investigated data from automatic bicycle counts in Germany and Austria. Albeit the existence of Type-I-curves with a morning peak and Type-II-curves with an afternoon peak (Figure 3.5 a), the majority of biking facilities has a Type-III-curve with peaks both in the morning and evening during working days. Saturdays and Sundays have only an afternoon peak, with a time-shift similar to road traffic. Also the hour-by-hour silhouette is similar to road traffic (Figure 3.5 b), albeit bike traffic stronger depends on weather and light conditions. The land use and the connected trip purposes in the vicinity of the counter strongly influence the load curve, since transiting flows are comparatively less influential due to the short trip length of biking.

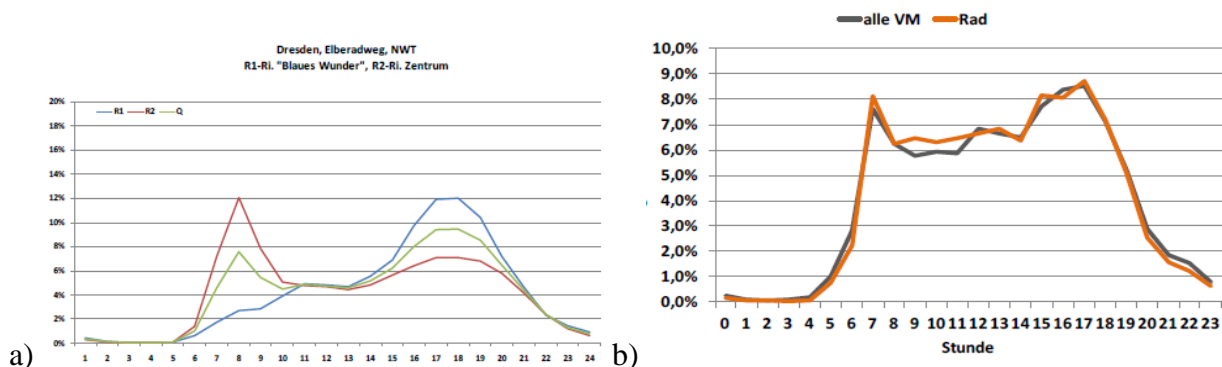


Figure 3.5: Bike load curves in Dresden (a) and from the survey MiD2008 (b) [Schiller et al., 2011].

Chapter 3 of the HCM [TRB, 2010] summarises similar findings from Portland (USA) and Copenhagen (Denmark) for bike traffic. The also therein contained pedestrian load curve from one sample cannot be generalized and is therefore not depicted here.

### Classification of Real Load Curves

The load curve types presented before generalise the development of real-world traffic over time. In the following, real-world load curves are investigated. This action was performed to cross-check the standard load curves and to reveal any peculiarities, if existing.

The real-world measurements stem from different projects performed in the past. The ones discussed in the following are summarized in Table 3.2. The table lists the number of detectors within the regarded area as well as the number of days the data set contains (“Covered Time”). Because most real-world measurements contain errors, only those detectors that were correct over a complete day were used. The respective number is given in the table as “Used”, denoting the number of complete time lines over a day. Please note that besides broken time lines, also time lines from week end days and Fridays were removed. Mondays, albeit being known to be different from Tuesday-Thursday weekday type, were kept unless differently reported in the following evaluations. The table’s “Single Lane” column indicates whether the respective detectors cover a single lane (“x”) or multiple lanes in one direction of a road cross section (“-”). Finally, the table gives the time resolution of the detectors.

**Table 3.2: Summaries of the data sets used for load curve analyses.**

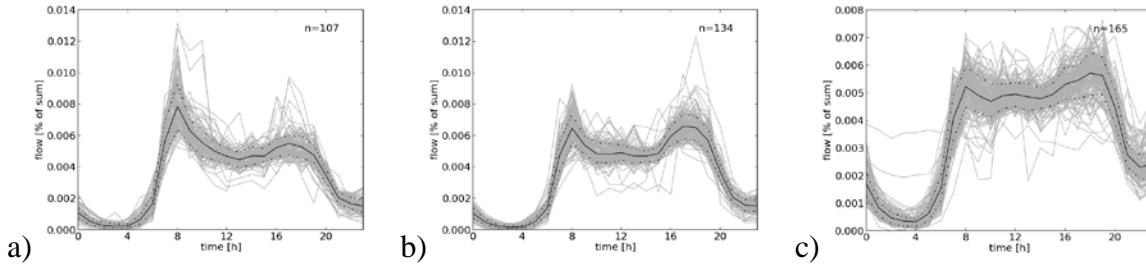
Name	Number of Detectors	Covered Time	Used	Single lane	Time Resolution
Bologna	638	3 days	407 / 1914	-	5 min
	The Bologna data set was supplied by this city’s communality within the iTETRIS project. The measurements contain the information about the number of passed vehicles and additionally a quality index ranging from 0 to 100. For the subsequent evaluations, measures which have a quality index of 100 over the complete day were used.				
Cologne	75	6 days	37 / 454	x	5 min
	The Cologne data set was supplied for DLR-internal projects DELPHI and VABENE by the municipality of Cologne. For the subsequent evaluations, only inductive loop measures were used, other sensors were neglected due to having different sensing frequencies (30 min and 1 hour). In addition, detectors that were valid for less than 280 intervals (of 5 min) were neglected.				
Brunswick	522	2 months	2722 / 8251	x	1 min
	The Brunswick data set was supplied for DLR-internal project AIM by Bellis AG who operates Brunswick’s traffic management facilities. For the subsequent evaluations, detectors that were valid for less than 1300 intervals (of 1 min) were neglected.				

The given measures were resampled to a frequency of 1 hour. Then, the so obtained curves were normalised by their respective maximum and were clustered afterwards, using the *fastcluster* package for Python [Müllner, 2013] with the Ward metric. For each data set, the respective time lines were joined into two to fifty clusters. The obtained classifications were investigated visually, first.

In a second step, the average of the load curves for a single cluster was compared to the three major load curves from RWS. In the following, the results are discussed.

## Bologna

As shown in Figure 3.6, the clusters obtained from the Bologna data set separate three load curve types<sup>5</sup>. The original measures are given in grey. The mean of all members is given as a solid black line, while two stroked lines represent the 15 and 85 percentiles. The clean separation can as well be seen in the according similarity dendrogram that is included in Appendix B. Therefore and by investigating the results visually, other cluster sizes do not give any further insight.

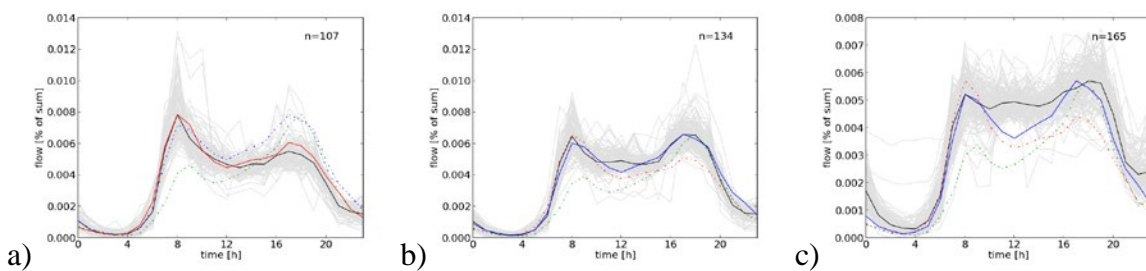


**Figure 3.6: Daily load curves from Bologna, classified into four clusters (see text for colour explanation).**

In a subsequent step, it was verified whether the seen three classes really match the ones given in the RWS. As shown in Figure 3.7, none of the clusters' average load curve is matching the afternoon peak. The colour and line style coding, used in subsequent images of this type is as following:

- solid black line: average of the cluster,
- red: morning peak type,
- green: afternoon peak type,
- blue: two peaks type,
- dotted lines: a RWS class that does not match the curve,
- solid lines: matching one.

It is to emphasize that the given Bologna time lines were shifted by exactly one hour to match the RWS types. The circumstance of later peak hours in Bologna – the morning one is between 8:00 and 9:00 while taking place between 7:00 and 8:00 in Germany – was already communicated by the municipality of Bologna during the iTETRIS project.



**Figure 3.7: Assignment of clusters shown in Figure 3.6 to the RWS classes.**

## Cologne

The Cologne data set shows two outliers, not shown in the following. In contrary to the clusters obtained from Bologna measurements, the remaining clusters found in Cologne measurements match all three RWS load curves, as shown in Figure 3.8.

<sup>5</sup> Albeit broken detectors were removed a-priori based on rules named in the text, some further measurement time lines were found that seemed to be broken. Usually, they separate clearly from the remaining clusters, yielding in clusters with only one element. They are not shown herein.

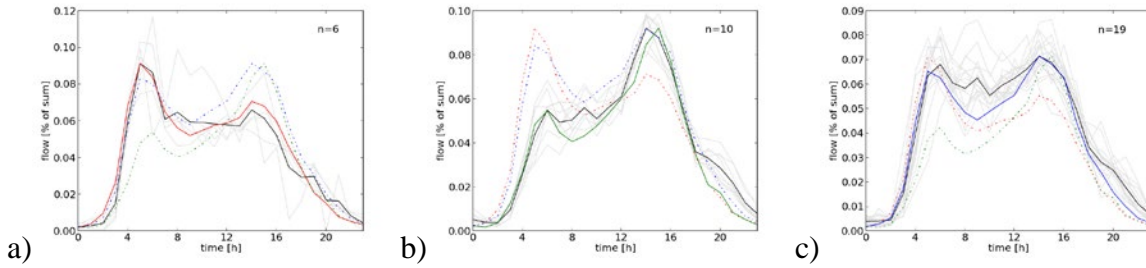


Figure 3.8: Assignment of Cologne clusters to the RWS classes.

## Brunswick

The Brunswick data show several peculiarities that are already visible within the similarity dendrogram (see Appendix B). On the top level, one can find two classes, each covering about a half of the detectors, both shown in Figure 3.9. Both have a “plateau”-like shape and differ in the magnitude of the traffic flow during the daytime.

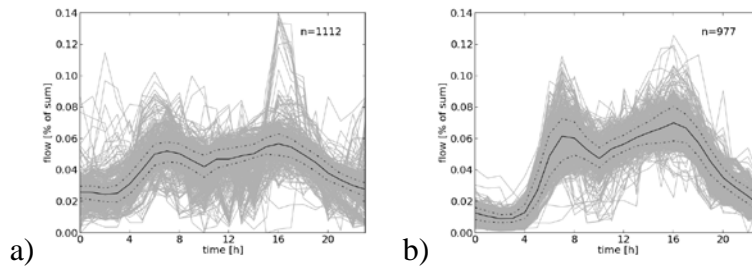


Figure 3.9: Daily load curves from Brunswick, classified into two clusters.

When going deeper, the standard classes appear (see Figure 3.10 a-c), albeit their average evolution is assigned to the “two peaks” RWS curve only. In addition, one can find a relevant amount of shapes where the demand goes back in the middle of the day to the low level one can usually find during the nights (Figure 3.10 d, e).

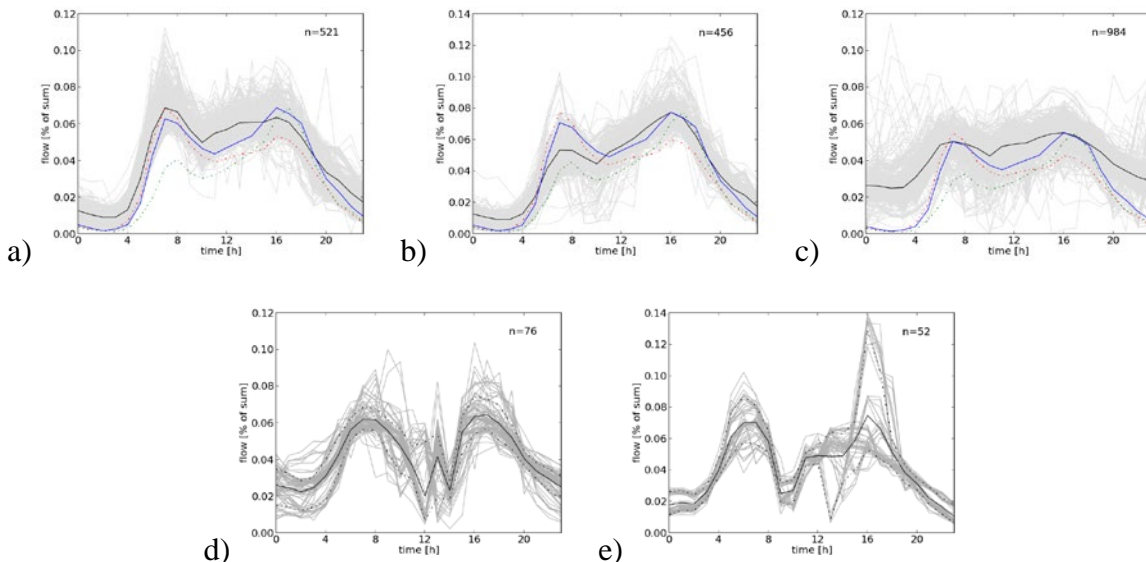


Figure 3.10: Daily load curves from Brunswick, classified into five clusters; top: with assignment to RWS classes.

It should be noted that for the investigation the time lines from Mondays were dismissed, because they contained shapes similar to the Sunday shapes shown in Figure 3.1. Regarding the overall unconformity of the results, the input data as well as their processing chain should be revalidated.



## Summary

The classification into three major clusters could be partially confirmed. Several special cases were found, some specific for a city. But in general, the investigated data matches the evolution of the daily load curve types defined in the RWS well (see Figure 3.7 or Figure 3.8).

The reduction to three major curves allows simulating the most common combinations of flow time lines over the day in reasonable time and additionally simplifies understanding of obtained results. Therefore, and because of the good matching between the RWS curves and real-world data, the decision to use the RWS curve types for evaluating traffic lights seems to be well motivated. The reconstruction of daily curves using multiple sinus waves as described in [COLOMBO D1.1, 2014]<sup>6</sup> seems to be less useful.

## Stochastic Approach

The traffic volume of a particular hour within the same category of days, i.e., work days, holiday work day, and Sunday, underlies a stochastic variability due to traffic flow processes and individual decisions of traffic participants [Lämmer, 2014]. These changes can be described with the variation coefficient  $c_h$  [%] for each hour. Figure 3.11 shows results of an investigation where this coefficient differs from hour to hour, with being higher (>5%) at hours of low traffic volumes, i.e., between 9 pm and 5 am, than during the day (<5%). This fluctuation is somehow controversial with the load curve type approach further up which assumes the same ADT-factor for all hours of the day. In this case the variation coefficients of each hour of the day are equal, and also equal to the variation coefficient within the respective day category of the year.

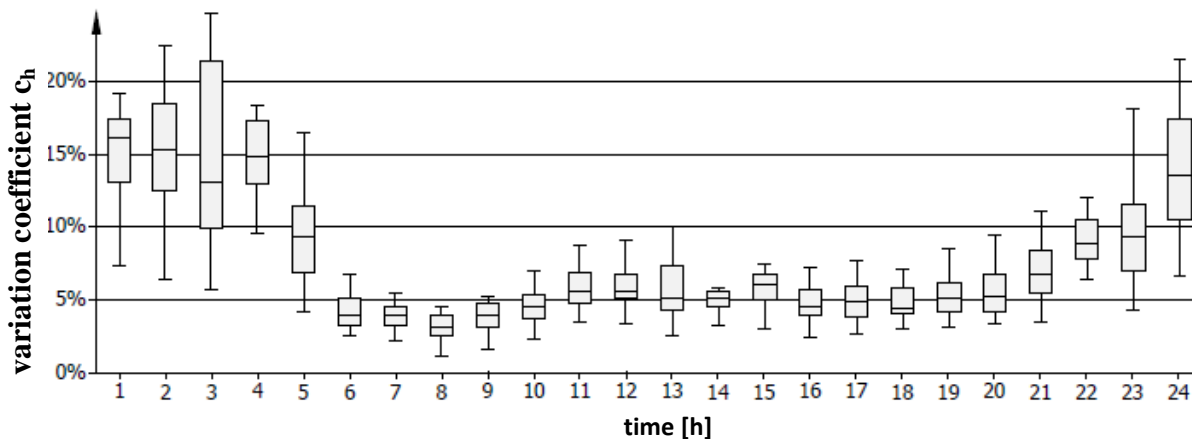


Figure 3.11: Variation coefficients  $c_h$  throughout the day Tue-Thu in Dresden [Lämmer, 2014].

## 3.2 Short-Term Traffic Flow Variability

### Statistical Approach

The discussed daily load curves usually describe the development of traffic amounts at a time scale of one hour. The traffic flow within one hour underlies a stochastic variability due to random user demand as well as traffic control processes. For short time intervals  $\Delta t$  between 1 hour and 5 minutes a Gaussian / Normal distribution of its traffic demand  $q_{\Delta t}$  can be assumed [Lämmer, 2014], for which [Axhausen, 2014] gives the following equation

$$q_{\Delta t} = \frac{q_h}{n} \times f, \quad (3-1)$$

where

<sup>6</sup> section 4.3.2 “Generating synthetic Demands”

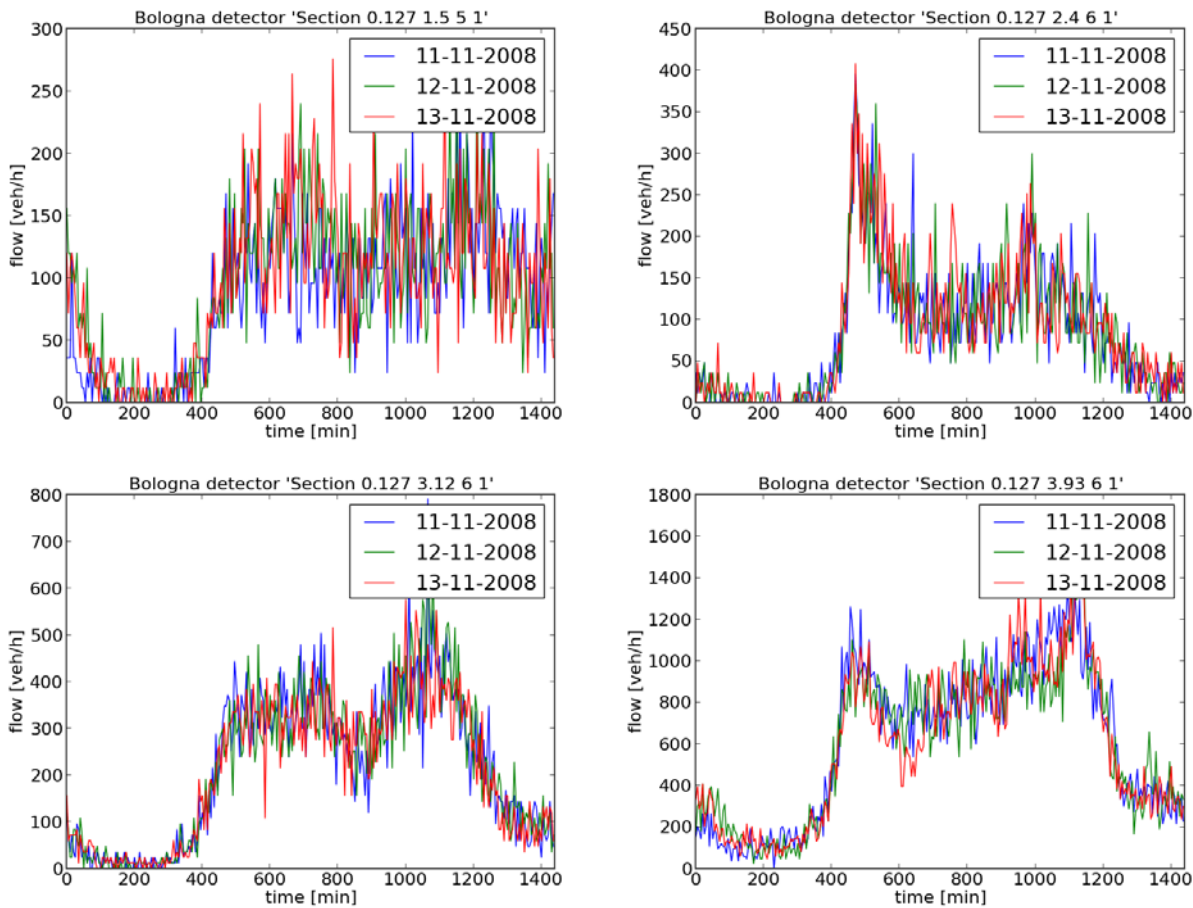
$q_h$ : traffic demand of the whole hour [veh/h],  
 $n$ : amount of intervals within one hour [-], and  
 $f$ : normal distributed factor with  $N(\bar{f} = 1, s = \frac{4}{\sqrt{q_h}})$ .

According to both references, aggregation time intervals smaller than 5 minutes incur strong influences from up-stream platoon formations at traffic light controlled intersections, and therefore include patterns of the cycle time.

Chapter 4 of the HCM [TRB, 2010] has a formula to calculate the peak hour factor (PHF) from 15-minute-intervals.

### Investigations

When investigating measures from the real-life, strong fluctuations at small time scales may be observed, even if the overall shape of the load curve is preserved. Figure 3.14 shows this by example for three consecutive working days (Tuesday-Thursday) of flow measures collected by four arbitrary chosen inductive loops located in Bologna.

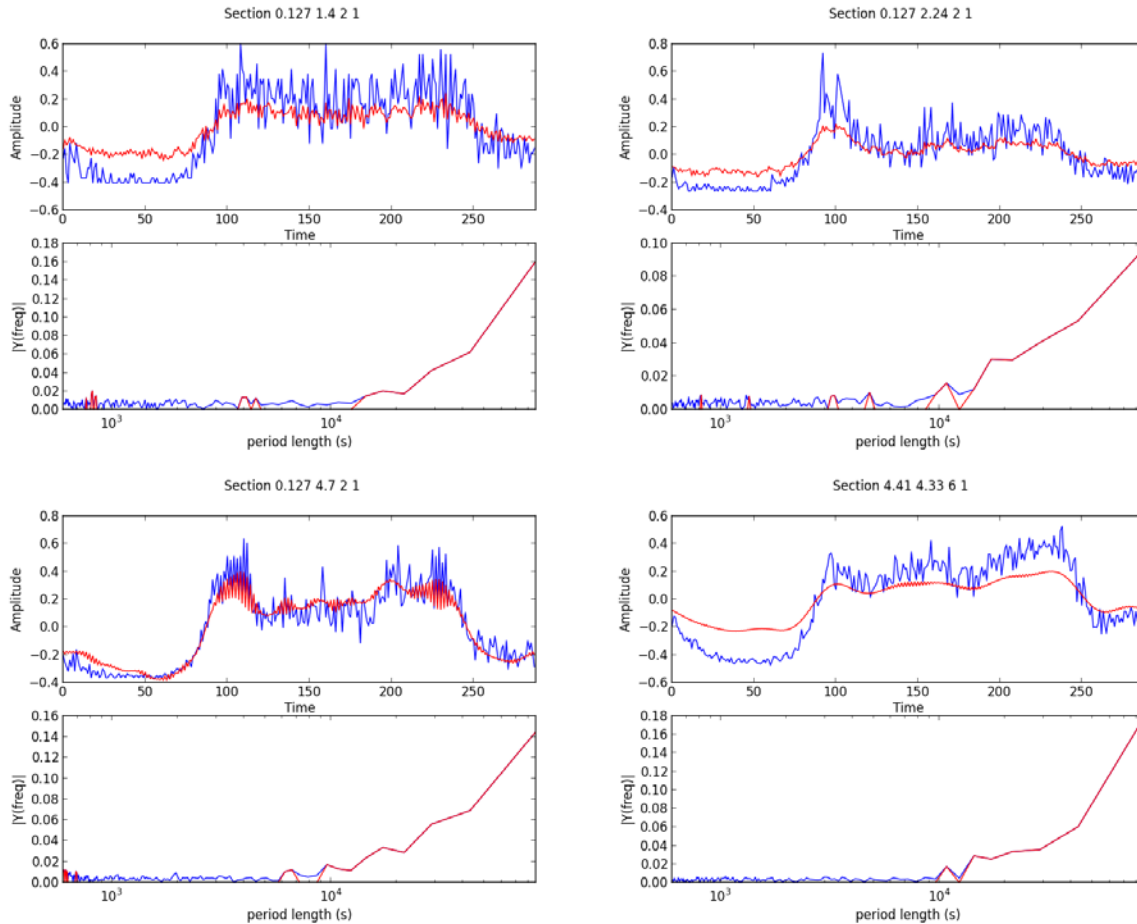


**Figure 3.12: Examples of real-world flows with a resolution of 5 min (source: Bologna traffic data).**

From the viewpoint of simulating and evaluating traffic lights, the frequencies of traffic arriving at an intersection are important, as a traffic light might adapt to certain frequencies only. In principle, this is already the case: actuated traffic lights react on inter-vehicle distances, traffic light algorithms that look at lane or edge areas, such as the self-organising traffic lights developed in COLOMBO, react on changes in the demand in magnitudes of minutes, while weekly switch plans consider changing traffic flow levels over a day by different pre-timed signal plans. To determine

the arrival frequencies, a discrete Fourier transformation (DFT) was performed on the available inductive loop measures that were presented before.

The Bologna data, originally sampled in intervals of 5 min, were investigated, first. On the original data, the DFT was performed using the *numpy* Python package. From the obtained frequency amplitudes, those above 1 % were chosen and were used to reconstruct the demand, the other frequencies were dismissed. The following figures show detector measurements and frequencies as following: in the top part, the original detector vehicle flow measures aggregated into intervals of five minutes are shown in blue. The demand reconstructed from the major frequencies is shown in red. The lower part of the figures shows the frequencies' amplitudes by the period length (in [s]). Those that are above the chosen threshold are shown in red, the ones below in blue. Some examples from the Bologna data set aggregated over 5 min intervals are shown in Figure 3.15.

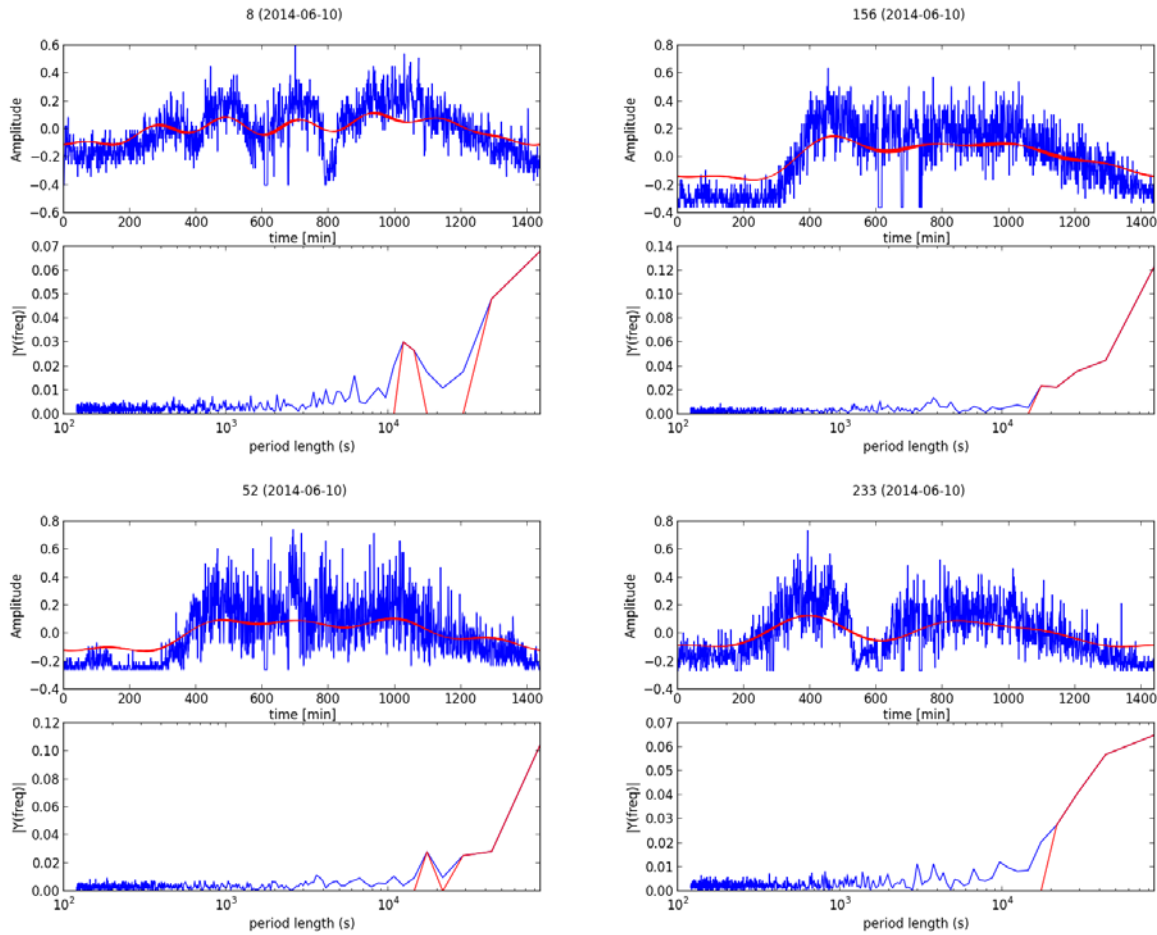


**Figure 3.13: Examples for frequency distributions found in a daily load curves obtained from detectors located in Bologna**

Several observations can be made. First, the lower frequencies, representing longer times, have higher magnitudes. This resembles the big changes in traffic flow over a day (“load curves” discussed before). Still, they are mostly not sufficient to reproduce the overall demand curve. The reason therefore is that the steepness of changes is rather similar to a square wave that can be approximated using a large (infinite, in fact) number of frequencies only. Second, one can hardly find peaks in higher frequencies. This means that no major fluctuation frequencies exist to which a traffic light would have to adapt to at lower time scales.

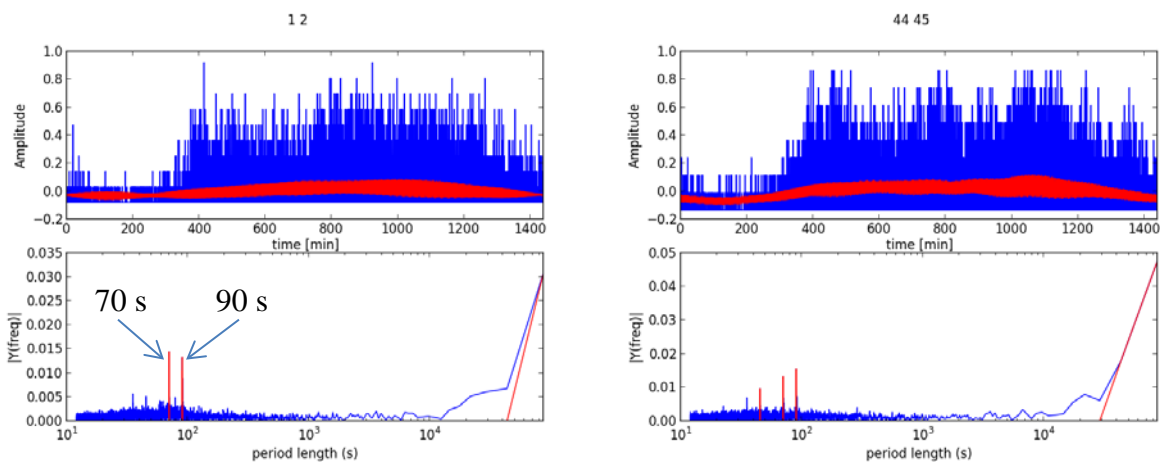
These observations hold for almost all of the investigated detectors. They are as well correct when investigating data from the other cities named before for aggregations between 1 min and 30 min. Some examples are shown in Figure 3.16. They use the same representation as used in Figure 3.15.





**Figure 3.14: Examples for frequency distributions found in data aggregated over 1 min; top: detectors from Brunswick, bottom: detectors from Nuremberg**

But when going to a lower aggregation interval of .1 min (6 s), one can find the above described cycle-time-induced structures as shown in Figure 3.17 with data from the DLR Berlin test track. It should be noticed that such an aggregation interval is not used in reality. Here, it was obtained by sampling the not aggregated, single vehicle passes.



**Figure 3.15: Examples for frequency distributions found in data aggregated over .1 min (6 s) from detectors located at the DLR Berlin test track**

In Figure 3.17 several peaks at higher frequencies can be clearly seen. The major ones are located at frequencies of 70 seconds and 90 seconds (see Figure). The one located at 90 s is assumed to resemble the influence of the upstream traffic light which cycle time is known to be 90 s. The other

ones are assumed to be generated by vehicles that turn from other directions into the road the detector is placed at. Two further, slightly lower peaks exist at 35 s and 45 s in some of the time lines. They are assumed to be side-effects of the previously discussed, major peaks. A validation using a simulation model was not performed.

The results of analysing the frequencies have two implications: first, examining traffic flow fluctuations at lower time scales is possible as demonstrated using data aggregated in bins of .1 min. Second, the major found influence that generates flow fluctuations is a traffic light. No other patterns could be observed.

### 3.3 Turning Ratios

Each approaching traffic stream on an intersection normally divides into vehicles turning left, right, and straight, if not particular turns are forbidden. If the resulting turn(ing) ratios underlie any patterns or correlations with the time of day, type of day, i.e., working day vs. weekend, or the overall traffic volume and hence importance of the respective approach is hardly researched. To get a first idea lane count data of three intersections on an urban dual carriageway in Helmond/NL were investigated during four subsequent weeks. As some lanes serve mixed turns (straight+right) not all approaches could be included in these investigations. Time aggregation intervals are 60 minutes. Tuesdays, Wednesdays, and Thursdays were found to be similar and thus grouped into one day type (“TueThu”), while the other 4 days of the week were kept apart. Nevertheless at least Mondays and Fridays still resemble this type “TueThu”.

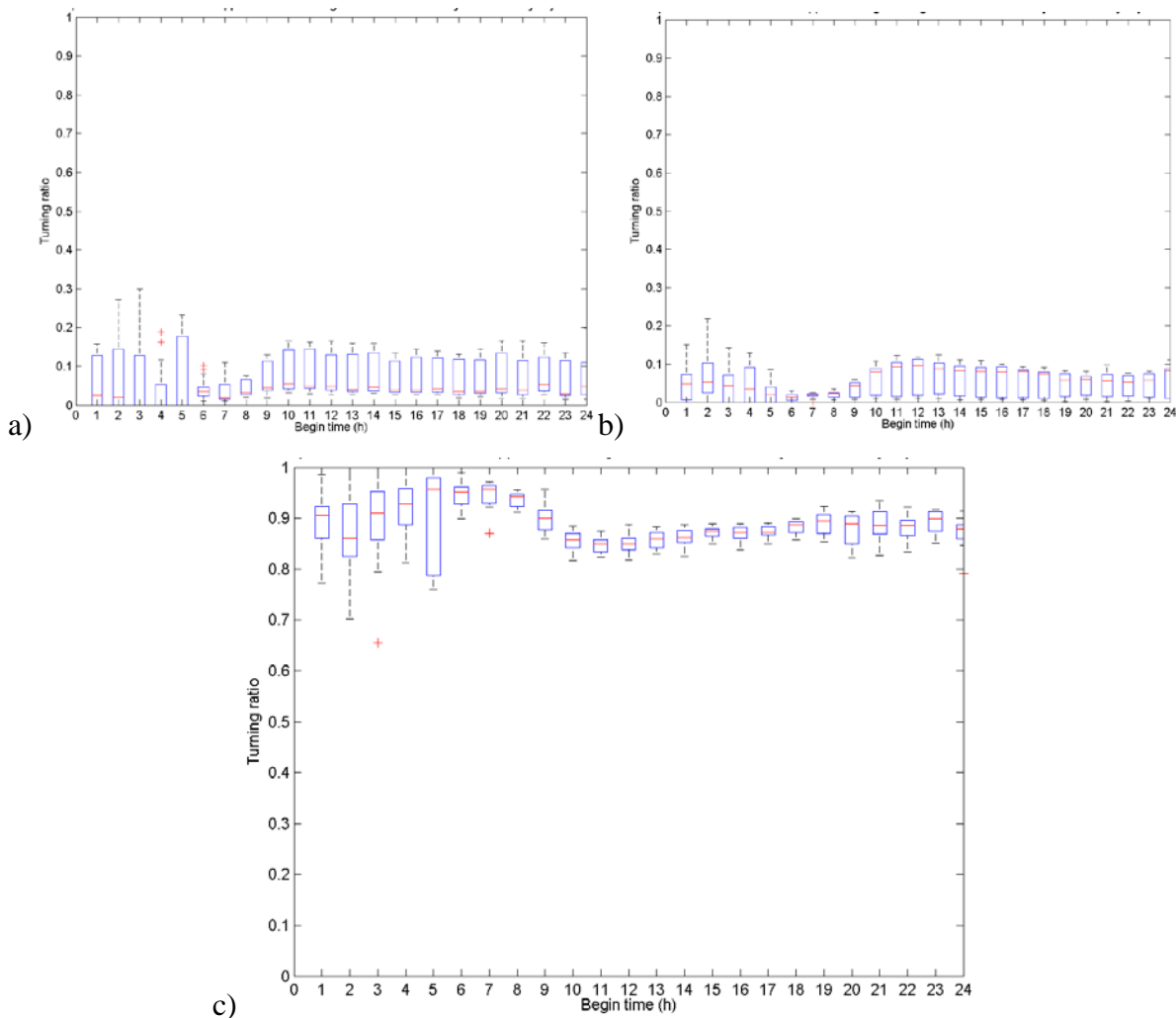
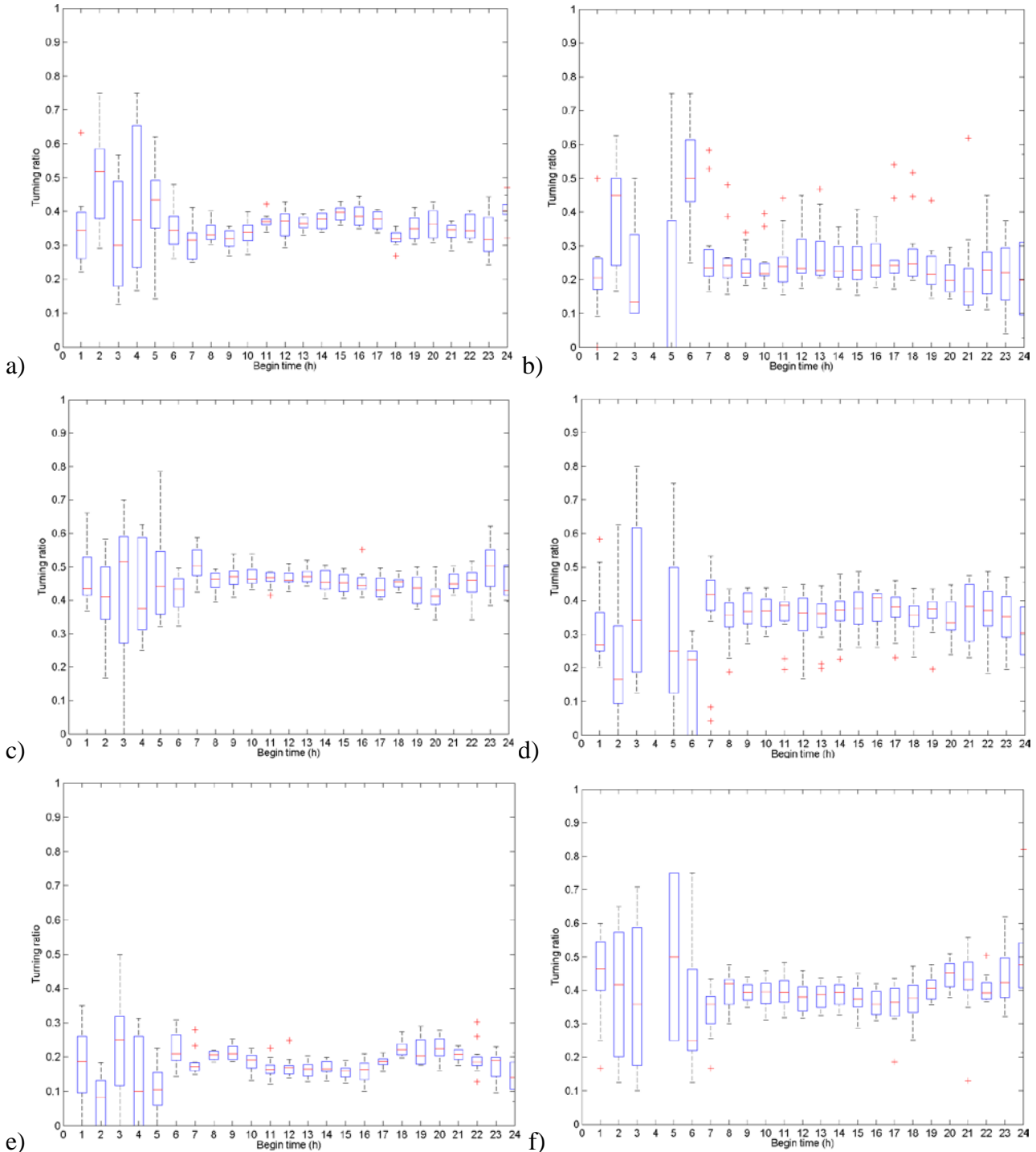


Figure 3.16: Variations in ratios for a) left, b) right, and c) straight turns of both major approaches throughout a day.

At first, both major approaches of intersection 102 are presented in detail. The main direction at both is straight. As can be seen in the boxplots Figure 3.18 night hours have a higher variation (and variation coefficient), a similarity to the traffic volume investigation in Figure 3.13. But in contrary to changing traffic volumes throughout a day, expressed by daily load curves, the turning ratios remain relatively constant within and between time intervals, with the straight turn share over 85% due to being the main direction. Furthermore both approaches seem to resemble each other in their turn properties, since the boxes are rather narrow.



**Figure 3.17: Variations in ratios for a+b) left, c+d) right, and e+f) straight turns of the Northern (a, c, e) and Southern (b, d, f) minor approaches throughout a day.**

The minor approaches in Figure 3.19 with a lower absolute traffic volume show a unpredictable behaviour during night hours. During daytime they have a smaller variation within each time

interval but a slightly higher variation between the time intervals than the major approaches. Furthermore there is no clear main direction.

Putting all major approaches of all intersections together would be the next step of data mining but yields so far no common picture anymore.

### **3.4 Vehicle Fleet Compositions**

Within the (motorised) vehicle traffic flow different kinds of vehicle categories, e.g., passenger car or truck, as well as different engine types (fuel, emission standard, gears, and driver support) can be found. The distribution of these vehicles within a certain fleet is different, considering spatio-temporal circumstances like the country or region, the land-use around a certain junction, the type of a road, or the year. This counts for the share in driven kilometrage accordingly, due to different usage patterns of, e.g., private and commercial vehicles. The resulting distributions influence driving dynamics like slower acceleration of trucks after a stop, as well as the emissions.

Traffic counts and detectors deliver at least separate counting numbers for passenger cars and trucks (including buses), but can diversify up to 8+1 classes (cf. German TLS [BMVBS, 2012] Appendix 2). The topic of pollutant emission is covered within COLOMBO by the work package 4 “Optimisation of Energy Consumption and Emissions”. More details on the emission classes can be therefore found in COLOMBO’s deliverables D4.2 and D4.3.

## 4 New Methodology for Evaluating Traffic Light Systems

As could be seen in Chapter 2 neither a common nor a sophisticated methodology for TLC evaluation exists. This starts already at the selection of performance indicators (PIs). As presented in [COLOMBO D1.1, 2013], a lot of different PIs exist. Using them all and summing up will not necessarily indicate how well a TLC strategy performs. For instance, which of the following control strategies is better? One that has an average delay time of 90s, while the maximum delay time is 500 s or a second strategy that has an average of 100 s but a maximum of only 120 s? Similarly, also the importance of different PIs, such as emissions and delay time, has to be weighted and this basically holds for all measurements. This means that uni- and multivariate statistics have to be applied onto the measurements for the evaluation.

Arriving at one number is important, not only to have a final “grade” of a traffic control strategy, but also to be suitable for an automatic tuning and configuration algorithm [Stüttgen and López-Ibáñez, 2013], like performed in COLOMBO’s WP 3. Automatic configuration requires only one criterion, as with multiple criteria an algorithm cannot determine which test run was best, unless all criteria are best for one scenario.

Within this chapter, an attempt to derive such a single measure is presented and discussed. The presented measure joins several objectives of a TLC performance, weighting them according to the goals of the traffic management and the city’s traffic policy behind. In addition, it allows an automatic optimisation as outlined and therefore helps COLOMBO WP3 in achieving its tasks. This chapter will first discuss which performance indicators are selected for the evaluation (4.1), followed by the development of a single measure with presenting the assignments of weights to the PIs, and the transformation rules (4.2).

### 4.1 Performance Indicator Selection

Stakeholders potentially affected by a TLC are the road users, the road operator, and to a lower degree the residents. The main objectives, as identified in [COLOMBO D1.1, 2014], differ between these stakeholders. For road users are mobility, resource efficiency, and user satisfaction. For the road operator these are mobility and environmental impact.

These measurements should be acquired accurately and in an unambiguous, reproducible way as described in [Blokpoel et al., 2010]. An overview of criteria and PIs was already presented in [COLOMBO D1.1, 2013]. An overview of criteria/PIs was already presented in D1.1. This resulted in a list of Performance Indicators (PI) with formulas and methods how to calculate them. The list of resulting PIs on an intersection level, together with the rationale to change, discard or adopt for evaluation purposes, is shown in Table 4.1. The goal of the PI selection is to acquire a set of indicators that is as complete as possible with minimal overlap. Most interesting are the choices made regarding waiting time, delay time, stops, and CO<sub>2</sub> emissions. Since delay time is basically the sum of waiting time and acceleration and deceleration losses, there is no value in having waiting time when delay time is already used. The amount of stops is a very common performance indicator as it is an indication for both emissions and driver comfort. In some cases when waiting time was used instead of delay time, it would compensate for the acceleration and deceleration losses. However, with the introduction of more accurate emission models like PHEM, this factor is not necessary anymore. The CO<sub>2</sub> emission indicator also almost linearly overlaps with other pollutant emissions and fuel consumption and is therefore a good general emission indicator. This is presented in detail in COLOMBO’s deliverable D4.3 “Pollutant Emission Models and Optimisation”.

Table 4.1: PIs used for evaluating traffic lights.

Criteria / PI	Definition / Comments
<b>Objective: Mobility</b>	
maximum queue length	Will be used as a constraint criterion*
mean queue length in front of the junction	Will not be used
total waiting time at intersection	Discarded, overlaps with delay time
mean waiting time in front of the intersection	Discarded, overlaps with delay time
Demand waiting time	This is the waiting time for the first vehicle in the queue, not considering double stops. Long waiting times for the first vehicle in a queue increases the risk of red light violation. In literature often the cycle time is used for this, which places an upper bound on the demand waiting time. The demand waiting time also takes situations into account when a stage is skipped during a cycle because there is no demand or when the first vehicle arrives long after the start of red. This indicator is used as a constraint criterion.*
mean delay per signal group	Will be used
Standard deviation of delay time	New: The predictability of the delay time is an important factor, as this enables drivers to plan their trips more efficiently.
number of stops	Discarded, overlaps with delay time, CO <sub>2</sub> emissions and driver perception.
Latent demand	Difference between total demand and the amount of vehicles which managed to enter the network.
<b>Objective: Resource Efficiency</b>	
mean fuel consumption	Discarded, overlaps with CO <sub>2</sub> emissions
junction saturation (I/C-ratio)	Discarded, this is dependent of the traffic demand and not of the control strategy performance
<b>Objective: Environmental Impact</b>	
mean exhaust emissions for pollutant $x$	Only CO <sub>2</sub> will be used, others scale almost linearly with CO <sub>2</sub> .
mean noise emissions	Will not be used, traffic control has little influence on this, traffic volume and speed limit are the main factors for noise
<b>New Objective: User Satisfaction</b>	
User acceptance	Described in more detail in 2.3. Important to review the end users perception of the performance.

\* Constraint criteria are only important if the configured maximum value is exceeded. In that case the solution is disqualified.

Note that only PIs which can be influenced by the traffic light controller are considered in this research. Speed differences and traffic volume can for instance also be used as an indicator for safety, but are not directly influenced by the traffic light controller. Additionally, safety requirements are also stricter, for example exceeding a maximum queue length which causes a

queue to spill back onto a highway is a dangerous situation. Therefore, these indicators are implemented as constraints.

## 4.2 Development of a Single Measure

### *Weighting Assignment*

To arrive at a single measure all PIs are given a weight. The higher the weight the more important the PI is. This is not the absolute importance as it also considers the inter-dependencies between PIs. For example the user acceptance has a high correlation with the PIs on which it is based, i.e., delay time and number of stops. Therefore, the weight is relatively low. Additionally, the weights are considered per transport mode. This results in differences between the classes, as can be seen in Table 4.2. For instance user acceptance was only derived for vehicles and considers red waves and number of stops. These factors are not applicable for pedestrians as the queues in practice never get longer than can be handled in one green phase and green or red waves do not exist either. The same holds for bicycles, although there is a possibility to investigate the factors for user acceptance and green waves for bicycles in the future.

**Table 4.2: Weight (Importance) of the used measures**

Transport Mode	Criteria/PI	Type	Proposed Weight
Pedestrians	Delay Time	Average	8
		Standard Deviation	2
	Waiting Time	Maximum	Constraint
Bicycles	Delay Time	Average	8
		Standard deviation	2
	Waiting Time	Maximum	Constraint
Vehicles	User Acceptance	Average	1
	Delay Time	Average	5
		Standard deviation	1
	Waiting Time	Maximum	Constraint
	CO <sub>2</sub> emissions	Average	3
	Queue Length	Maximum	Constraint
Public Transport	User acceptance	Average	1
	Delay Time	Average	5*
		Standard deviation	1*
		Maximum	Constraint
	Waiting Time	Maximum	Constraint
	CO <sub>2</sub> emissions	Average	3
	Queue Length	Maximum	Constraint
Overall	Latent demand	Total	Constraint

\* multiplied by the occupancy, note that the weights will have to be renormalized accordingly.

Public transport is a special transport mode which has the possibility to add an extra constraint indicator for the delay time. The class is also often subdivided according to schedule adherence. Vehicles that are behind schedule and possibly have passengers that miss a connection at the



destination of the vehicle, often receive a much higher priority than vehicles that are ahead of schedule and possibly even have to wait at the next stop to regain schedule adherence. As indicated by the \* in Table 4.2, the delay time and its standard deviation require a special treatment. These should be multiplied by (an estimation of) the occupancy of the vehicle. This is in contrast to the emissions and user acceptance which are just for the whole public transport vehicle and its driver, independent of the occupancy.

As can be seen each transport mode has a sum of weights of 1.0, while the number of PIs can vary between them. Another important factor in evaluation is the traffic demand of each transport mode (cf. [Hunter, 2010] and [Pulugurtha and Kusam, 2011]), which is included in two ways. Firstly, the latent demand, this is the difference between the actual demand and the amount of vehicles that could be inserted into the network during simulation, which should always be 0 vehicles. Lastly, the travellers' throughput should be taken into account to weight each mode's overall grade with respect to the total volume, to arrive at the final grade. For the number of passenger car travellers and public transport passengers the throughput will be multiplied with the respective occupancy (default car occupancy is 1.3 passengers/vehicles). The latter is an objective method, for example if there is one signal group for cars with a volume of 770 vehicles in an hour (equally to 1.000 passengers) and a conflicting group for pedestrians with 100 in an hour, this results in the cars being 10 times as important when the proposed weights are considered. In theory this is not a problem, since the constraint criteria ensure an acceptable solution. However, when local policies require more importance for a certain category, the weights can be scaled up or even normalized for the throughput of the different categories.

### *Grade Assignment*

---

Grading is the transformation from the achieved PIs into grades and takes place according to the concept of a "Level of Service (LoS)" (cf. section 2.1). The base of the grading scale ranges from 1 to 6; PI values between these limits of the category are continuously interpolated and beyond the base of the grade scale extrapolated. That way an empty network can theoretically get a zero while a severely congested network which does not trigger a constraint criterion can have a grade higher than 6. Capping the grade and a stepwise transformation  $PI \rightarrow \text{grade}$  would result in loss of information and is therefore not preferred in newer publications [Smart et al., 2014]. A multi-criteria analysis uses a grade scale from 1 (100% achievement of defined objectives) to 0 (worst), but this would complicate the comparison between that method and the LoS used by traffic engineers, while mathematically the methods are the same. Additionally, a fixed maximum grade results in loss of information when the maximum is exceeded.

The grade definitions from literature for delay time or waiting time are not all the same, as shown in section 2.1. Therefore, they have been consolidated for normal vehicles and for pedestrians, cyclists and public transport combined. The grade criteria are summarized in Table 4.3. This resulted in stricter criteria for public transport, pedestrians and bicycles since the risk of red light violation is higher for pedestrians and bicycles, while for public transport the delay affects all passengers as well. The standard deviation takes as a reference the expected standard deviation when vehicles arrive according to a uniform distribution during a fixed time cycle corresponding to the delay time associated with the same grade. For example when the average delay is 20 seconds, there could be a uniform distribution between 0 and 40 seconds, which results in a standard deviation of 12 seconds according to general statistical theory.



**Table 4.3: Grades for delay time.**

Grade	Average delay vehicles	Average delay pedestrians, bicycles and PT	Standard deviation vehicles	Standard deviation pedestrians, bicycle, PT
1	15	12	9	7.2
interpolation	linear			
6	90	72	54	43.2

The PI “user acceptance” is not linear, but a second order polynomial. Proposed grades are shown in Table 4.4, together with the CO<sub>2</sub> emissions. A new method to determine the threshold values is described in the following section.

**Table 4.4: Grades for user acceptance and CO<sub>2</sub> emissions.**

Grade	User acceptance	CO <sub>2</sub> emissions (COLOMBO fleet)	CO <sub>2</sub> emissions (new vehicle fleet)
1	0.91	97 g/km	0.5 full stops + 10s
interpolation	$= -5.8UA^2 - .2UA + 6$	linear	
6	0	222 g/km	3.0 full stops + 60s

### *CO<sub>2</sub> Threshold Determination*

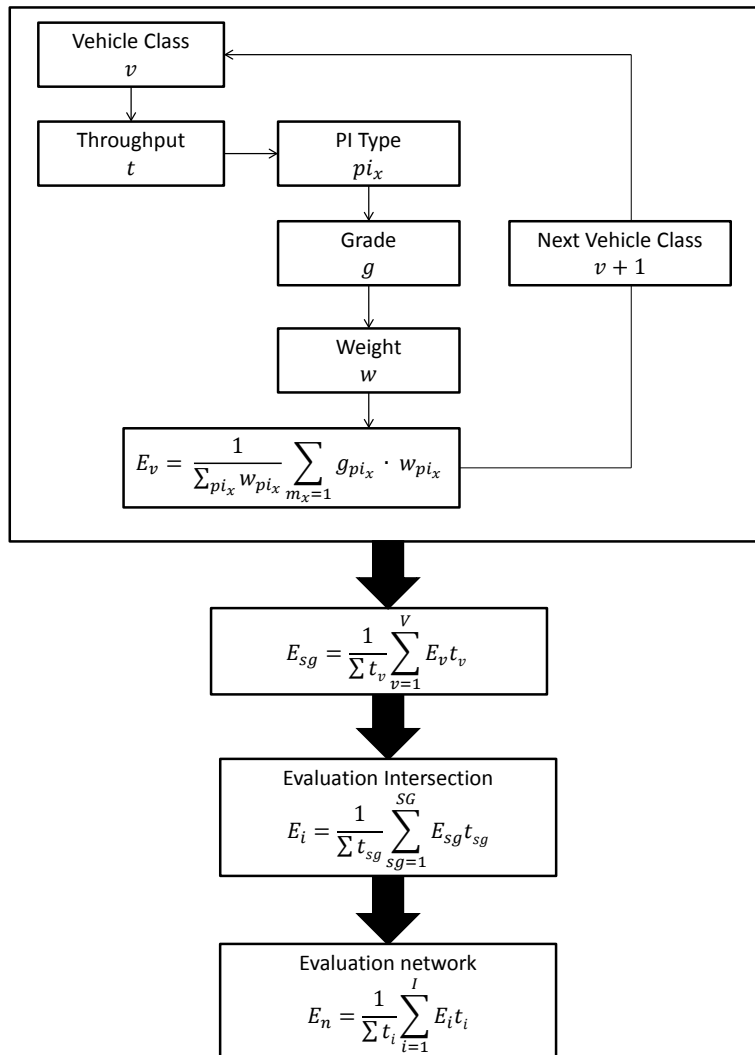
For CO<sub>2</sub> emissions the unit is taken in grams per kilometre as this is a very common unit also for the automobile industry. The per-kilometre-normalization also gives the advantage that it compensates for different distances between intersections. Otherwise longer distances would have a severe disadvantage in the scoring for total aggregated emissions. The starting point for the grading method is that both stopping and idling add to the emissions, while driving at a constant speed is the most efficient. The reference measures are determined by a virtual test track with a length of 500 meters which is a reasonable distance between intersections. On this test track the vehicles have to stop and reaccelerate 0.5 to 3.0 times on average while idling 10 to 60 seconds. With different fleet compositions or different speed limits these values have to be re-determined and therefore the rightmost column is included to describe the benchmarks of the procedure. For COLOMBO they were determined using the PHEM distributions for Austria of 2014 and a speed limit of 50 km/h. This PHEM distribution consisted of 3 main classes of cars, trucks and busses and 51 subclasses related to the subtype and emission characteristics, like for example Euro5 Diesel and Euro6 gasoline. Also note that specific vehicle classes with a different vehicle composition, like public transport, should follow a table with different values for CO<sub>2</sub> emissions. The resulting values for the emissions in Table 4.4, which range between 97 and 222 g/km, can intuitively be compared to the fact that vehicle emissions according to the New European Driving Cycle (NEDC) are usually higher for the urban part than for the extra-urban part. Despite the lower speeds in the urban part, which should theoretically result in a lower emission, this is far offset by the 3 full stops and 49 seconds of idling per cycle of 195 seconds. Letting the vehicles drive at a constant speed of 50 km/h resulted in only 72 g/km CO<sub>2</sub> emissions, which is practically not possible for a traffic light controller.

### 4.3 Implementation

The PI computation was realised as an application written in the Java programming language. The application reads outputs generated by the traffic simulation SUMO and extracts the necessary information to compute the involved PIs.

The application reads the road network, SUMO's "emissions output" that includes the positions of all vehicles for all time steps, and the vehicles' routes file. The files to read are given to the script on the command line. Using this information, the application computes the PI as given in Figure 4.1. The generated in-between results as well as the PI are then written as comma separated values (.csv)-file that can be read by other applications, such as Microsoft Excel or Open Office Calc.

This application can be called from the TLS Algorithm Evaluation System presented in this document's chapter 5.



**Figure 4.1: Structure chart of the single performance indicator computation algorithm.**

## 5 TLS Algorithm Evaluation System

The major goal of COLOMBO task 5.3 “Traffic Light Algorithms Evaluation Schema” was to deliver a software framework that benchmarks a new traffic light algorithm using a well-defined set of scenarios and performance indicators. The major task of this framework is benchmarking existing traffic light algorithms for comparing their performance – what was hardly possible up to now as shown in [COLOMBO D1.1, 2013]<sup>7</sup>. The second task is to help a traffic light algorithm developer to determine the strengths and weak points of her/his algorithm.

In this chapter, the realisation of COLOMBO’s Traffic Light Algorithm Evaluation System (abbreviated as “CTAES” in the following) is presented. One specific feature of CTAES is that it does not operate on single scenarios, but rather on what is called “scenario sets” in the following. A single scenario set contains several scenarios – where “scenario” denotes all the input files needed to perform a single simulation run including the road network, the traffic demand, and infrastructure settings. CTAES uses scenario sets for determining the development of a traffic light controls’ performance by iterating over ranges of one of the scenario’s parameters.

In the following, CTAES is presented, starting with the requirements put on it. Afterwards, Section 5.2 describes how the system was realised, focussing on the user experience. In Section 5.3, the concept of scenario sets is introduced and an overview about available scenario sets is given. The modelling of traffic participants is then given in Section 5.4, followed by a description about how the performance indicators are computed in Section 5.5. Section 5.6 gives some implementation details and Section 5.7 reports about some first experiences. The scenario sets are described in detail by an example application of comparing the fixed-cycle traffic light control against an actuated traffic light algorithm.

### 5.1 Requirements

The functional and non-functional requirements to the software that realises CTAES are listed in the following. They are differentiated into *must* and *shall* requirements.

#### Functional Requirements

---

##### Algorithms

- The user must be able to implement his own TLC algorithm.
- The system shall provide established algorithms (fixed, actuated) with which a base line scenario can be computed.

##### Scenarios

- The simulated scenarios must cover real-world characteristics of traffic.
- The simulated scenarios must allow to investigate the reaction of the investigated TLS algorithm when being confronted with a specific characteristic.
- The simulated scenarios shall include complex scenarios with irregular occurrences of special cases.
- The simulated scenarios must be compliant to TLC design guidelines and regulations.

##### Output

- The generated PIs must be well-defined
- The generated PIs must include the ones defined in D1.1
- The system shall supply vehicle trajectories which include vehicle type, speed, street/edge ID

---

<sup>7</sup> section 2.6 “Scientific Literature”

- The system shall supply a network topology which includes the locations of the signal groups to which the vehicle trajectories can be mapped.
- The system shall supply the traffic demand as a total demand per signal group.
- The system shall calculate the performance indicators which are listed in Table 4.2.
- The spatio-temporal aggregation level as well as the aggregation level for vehicle classes or any other scenario parameter should be definable by the user. E.g.:
  - spatial: lane; turn; edge; junction; O-D-pair;
  - temporal: 15 minutes; hour; different signal program; day;
  - technology: equipped vehicles (type A; type B); passenger cars; HGV; electric vehicles;
- Aggregated results shall be given either for each existing object of the respective parameter, or for a freely edible selection by an object identifier. For spatial objects areas
- The system shall further synthesize the evaluation value by applying the evaluation procedure with its edited weights and relevant PIs.
- The results shall be printed by edible 2D and 3D figures.
- The results shall be mapped onto the network graphic.

### Non-functional Requirements

#### Portability

- the system shall be executable on different operating systems, mainly MS Windows and Linux;

#### Extensibility

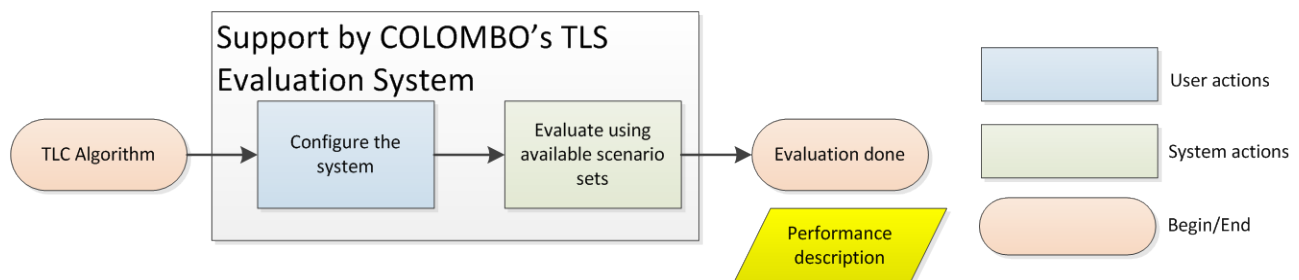
- the system shall be extensible by new scenarios;
- the system shall be extensible by new performance indicators;
- the system shall be usable with other traffic simulation software than the one used in COLOMBO;

#### Reproducibility

- The output results must be identic when an identic simulation is run again.

## 5.2 Realisation

The realised work flow for benchmarking a traffic light is as simple as it can be. Assuming that the traffic light logic to evaluate is already implemented in SUMO (like the case for COLOMBO self-organising traffic lights), the evaluation system may be directly started to obtain a large variety of PIs. The configuration step, depicted in Figure 5.1, is only necessary if additional non-default parameters have to be supplied to the tested traffic lights algorithm. In this case, a file that contains them must be generated<sup>8</sup>.



**Figure 5.1: The work flow for benchmarking a traffic light algorithm.**

<sup>8</sup> Assume, one has a single parameter “threshold” that shall be set to 1; the according file would then consist of a single line “threshold:1”.

As said, CTAES' major approach is to offer a set of well-defined "scenario sets" and evaluating an investigated traffic light algorithm's performance by executing those (or a subset of them) to obtain the performance indicators as defined in chapter 4.1 and listed in section 5.5. Therefore, the major functionality of the execution script reduces to determine which scenarios sets to execute, preparing them, performing them using the chosen simulation system, and extracting the resulting performance indicators (see Figure 5.1). The script offers some additional functionality to help the user in operating the database and/or selecting proper scenarios.

The responsible application (`ctaes_execute.py`) is set up as a command line application with no graphical interface. The user controls the behaviour of the script using command line options shown in Table 5.1. The system determines which scenarios shall be executed by evaluating the `--scenarios` option. Scenario sets to use have to be listed by name, here, divided by a semicolon (";"). Every listed scenario set may be configured using additionally given parameters.

**Table 5.1: Options for the scenario execution script.**

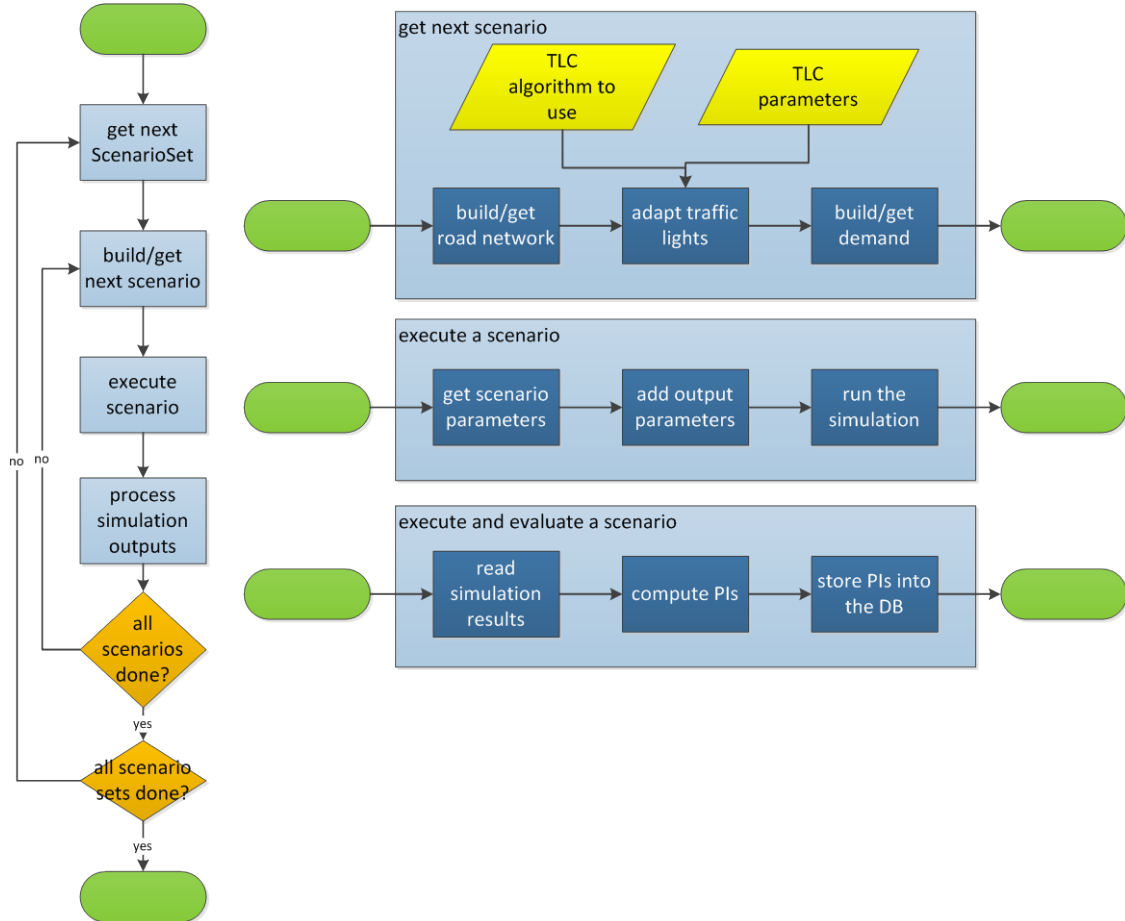
Option	Description
<b>Scenario Set Options</b>	
<code>--scenarios</code> <code>&lt;SCENARIOSET&gt;[;&lt;SCENARIOSET&gt;]+</code>	Defines the scenarios to execute
<code>--rebuild</code>	Rebuilds all scenario files even if they are up-to-date
<code>--sandbox-path &lt;PATH&gt;</code>	Defines where scenario files shall be stored
<code>--vtypes &lt;NAME&gt;</code>	The name of vehicle types definitions to use
<code>--sandbox-path &lt;PATH&gt;</code>	Defines where scenario files shall be stored
<b>Traffic Light Control Options</b>	
<code>--tls-type &lt;TLS_ALGORITHM_NAME&gt;</code>	Defines the traffic light algorithm to use
<code>--tls-params &lt;FILE&gt;</code>	Defines the file to read additional parameters from
<b>Control Options</b>	
<code>--runs &lt;INT&gt;</code>	Repetition number
<code>--simulator &lt;PATH&gt;</code>	Defines which executable shall be used for simulation
<code>--rebuild</code>	Rebuilds cached scenarios if set
<code>--extract-only</code>	If set, scenarios are built but not executed
<b>Database Options</b>	
<code>--db &lt;FILE&gt;</code>	Defines the database file to use
<code>--clear-db</code>	Removes all entries from the used database

The system builds the scenarios consecutively and adapts the definition of the traffic lights included to the algorithm that is specified using the `--tls-type` option. Additional parameters the algorithm may require may be read from a file named using the `--tls-params` option. This file

must contain key/value pairs that are embedded into the traffic light definition SUMO can read by the execution system.

The database file to use may be named using the `--db` option. If not existing, a new database file is built. An existing database may be cleaned by removing all entries using the `--clear-db` option.

As shown in Figure 5.2, the application iterates over the named scenario sets, then over the defined traffic light algorithms to evaluate, then over the scenarios of the current scenario set. The current scenario is built if needed or wished and its definitions of traffic lights are adapted to the needs of the traffic light control that is currently being evaluated. This process is performed  $n$  times where  $n$  may be specified using the `--runs` option. Such a repetition is wanted for obtaining statistically valid results.



**Figure 5.2: Steps performed by the system within the “Evaluate using a subset of available scenario sets” contained in Figure 5.1 and Figure 5.3.**

As building a single scenario may take some time itself, all files needed to describe a scenario (representations of the road network, the demand, and the infrastructure) are cached after being built and can be reused in later runs. As well, they can be used to rerun the simulation with different settings (e.g. other output options) to investigate the behaviour to a deeper degree. The path where the scenario files are stored is set using the `--sandbox-path` option, rebuilding the files can be disabled using the `--rebuild` option. As well, one may be interested in obtaining the scenario without running it. This may be achieved using the `--extract-only` option.

Besides references to all the files it consists of, a scenario is started with additional parameters that force the used simulation (SUMO) to generate certain output files. After the simulation run, these output files, which are usually stored in XML, are parsed to compute COLOMBO’s performance indicators. The resulting PI values are then stored in the used database under the current simulation

run's (scenario run's) ID. While section 5.5 describes the generation of COLOMBO PIs in detail, the database is described in section 5.6. As the plain values of performance indicators as well as their basic statistical properties are directly available within the database, the given evaluation tools concentrate on extracting and presenting them in an understandable way.

Most of the scenario sets iterate over some value ranges. The script “`ctaes_show_matrix.py`” shows the obtained PIs and/or their derivatives as a matrix. If more than one result for a simulation scenario is found in the database, the average value is used. The matrices of all processed algorithms are normalized by the found minimum and maximum values, allowing a direct visual comparison. The script is capable to retrieve and plot all computed PIs and their statistical properties. As well, it supports the generation of difference plots against a selected traffic light control algorithm. The presentation of the scenario sets in Appendix C uses this tool for visualisation.

Supporting a developer in designing a new algorithm is less straight forward than evaluating a given traffic light algorithm. At the current time, the system offers the following help:

- Reduce the granularity of iterations to get results fast
- Support the user in extracting further information about a simulation run than those used for benchmarking

Thereby, a complete work flow for evaluating a traffic light algorithm under development could be as shown in Figure 5.3. Please note that the user has to decide whether the quality of her/his TLC matches her/his expectations.

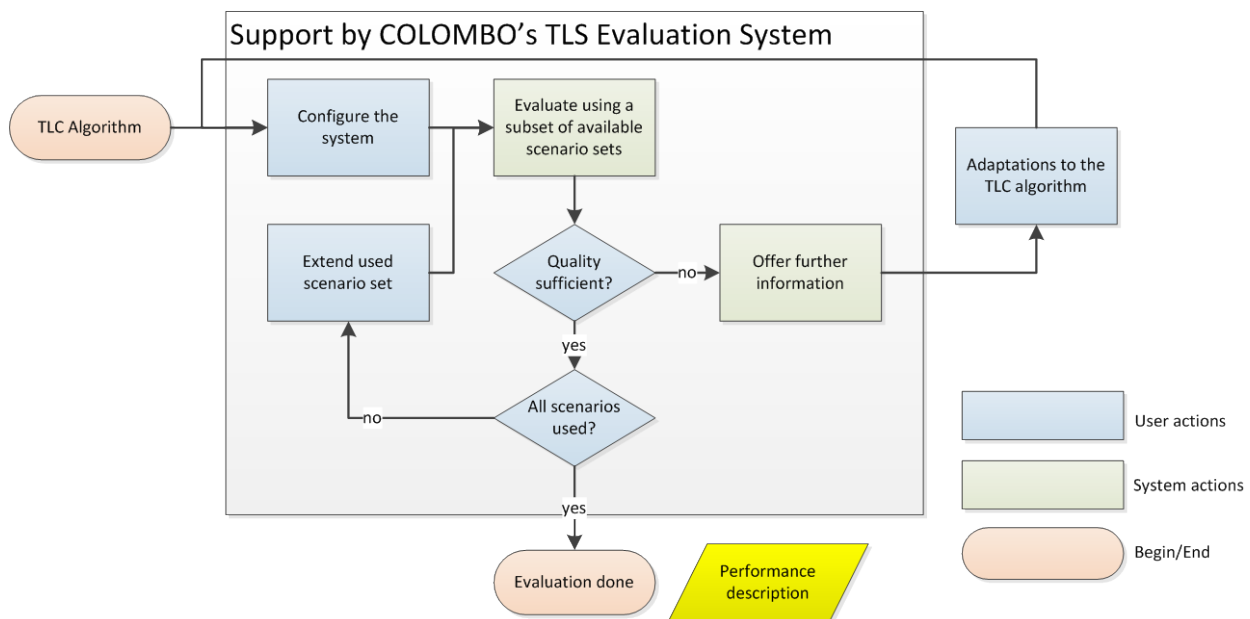


Figure 5.3: The work flow for improving a traffic light algorithm.

### 5.3 Scenario Sets

As stated in previous section, CTAES heavily relies on scenarios and scenario sets to fulfil its tasks. Every scenario set tests the performance of the examined traffic light control against exactly one of the exogenous parameters, which is iterated in meaningful and significantly differing steps. Some of these parameters could be adjusted to real-world traffic characteristics presented in chapter 3. All characteristics have been tried to be resembled by a set of scenario sets. Most of the thereby generated, artificial scenarios are not meant to resemble real-world situations. Instead, they are designed to evaluate a certain characteristic with least side-effects.



The major difficulty is the large number of possible combinations of the respective parameters; one may expect a traffic light to perform adaptation to:

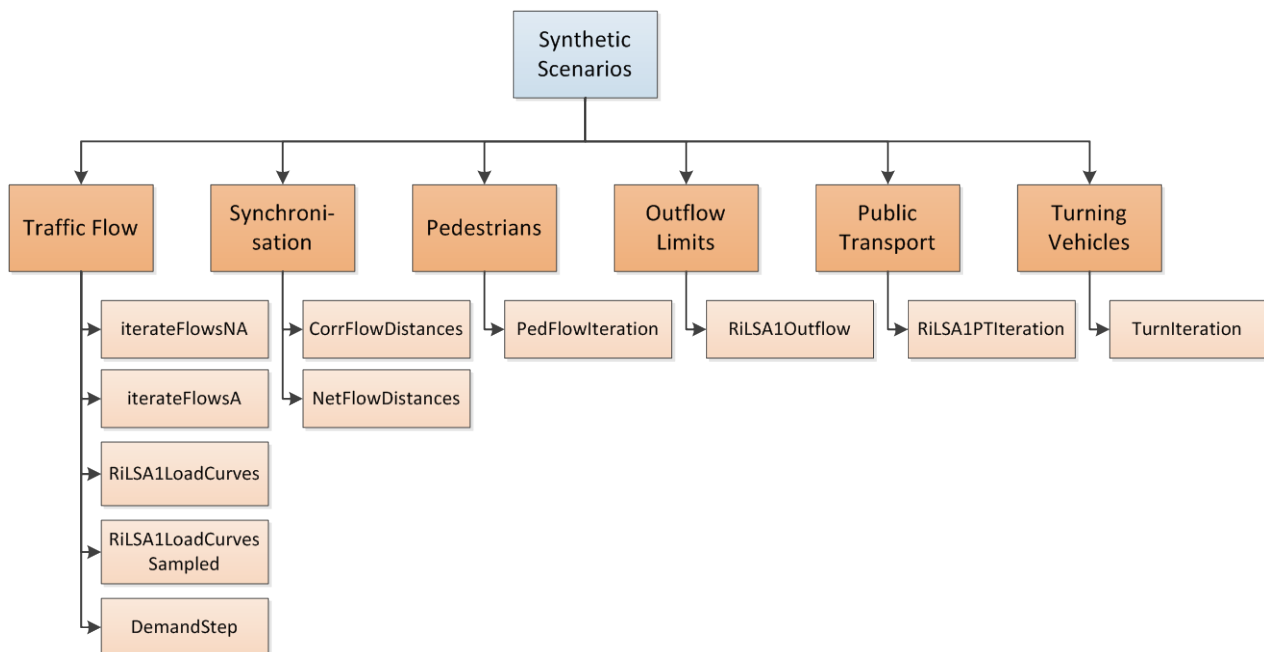
- changing (passenger) vehicle flows
- changing pedestrian flows
- changing bicycle flows
- approaching public transport
- approaching emergency vehicles
- approaching heavy duty vehicles

All these changes happen at different levels, as outlined in chapter 3, ranging from long-term plateaus over changes during a day to small-scale fluctuations. They may affect a single direction as well as all that run over the intersection. In addition, a traffic light control (TLC) may be sensitive to other aspects:

- limits on outgoing lanes that prevent vehicles from a single direction to cross the intersection
- high rates of turning percentages
- synchronisation of multiple traffic lights

Some of these influences cannot be modelled without taking into account other ones; at least a basic passenger vehicle flow is needed in all cases. When trying to combine all these influences, a tremendous number of scenario sets would become necessary. Additionally, not all of such obtained scenario sets could be tested for relevance at the current time, because traffic light algorithms that target at specific characteristics are neither known nor implemented. Some, such as regarding public transport of taking into account the limits on outgoing lanes are under development in COLOMBO, some others, such as dealing with changes in turn percentages are not.

To accommodate this issue, the following procedure was chosen: being the target of all traffic lights algorithms, the overall vehicle flow is covered in a large detail by five scenario sets. Specific aspects named before (e.g. public transport prioritisation) are covered by one scenario set, each. This is assumed to be sufficient for first investigations, but further scenario sets may be needed.



**Figure 5.4: Covering traffic light regulation aspects by the implemented scenario sets.**

It should be added that additional real-world scenarios are included, which usually cover a large set of aspects, albeit not treating them in an isolated way that removes any side-effects. The following



Table 5.2 summarizes the relationships between traffic characteristics and the currently implemented scenario sets. Please note that real-world scenarios are omitted here due to their heterogeneity. The scenario sets themselves are presented in a greater detail in Appendix C.

**Table 5.2: Coverage of discussed traffic characteristics by the developed scenario sets.**

	A: iterateFlowsNA	B: iterateFlowsA	C: RiLSA1LoadCurves	D: RiLSA1LoadCurves Sampled	E: DemandStep	F: TurnIteration	G: RiLSA1Outflow	H: PedFlowIteration	I: RiLSA1PTIteration	J: CorrFlowsDistancesA	K: NetFlowsDistancesA
static flow	x					x				x	x
flow fluctuations		x									
flow changes			x	x	x		x	x	x		
turning ratio						x					
outflow limits							x				
pedestrians								x			
public transport									x		
heavy duty vehicles			(1)	(1)			(1)	(1)	(1)		
synchronisation										x	x

<sup>(1)</sup> these scenarios have a pre-defined HDV amount

## 5.4 Traffic Participants

To measure a traffic light algorithm's performance, not only the road infrastructure and the traffic light logic itself have to be represented well, but also the traffic participants that cross the intersection(s). In the context of COLOMBO, two aspects have to be modelled properly, the dynamics of these participants, as well as their emission behaviour. Real-world traffic consists of different modes of transport, where each may be again subdivided into a high variety of different vehicles and/or behaviour variations. When trying to predict the performance of a traffic light for future scenarios, one should as well take into account changes in the participants' behaviours.

To simplify modelling of a scenario, usually a distinction between different participant types – modes of transport and/or carrier types – is done at the top level, first. The scenarios and scenario sets available so far include the following traffic participant types:

- passenger vehicle (“passenger”)
- heavy duty vehicle (“hdv”)
- bicycle (“bicycle”)
- pedestrian (“pedestrian”)
- (public) bus (“bus”)

Other, such as motorcycles or trams are not included in any of the modelled scenarios.

The used traffic simulation SUMO models vehicles and pedestrians. In SUMO, a “person” may use different modes of transport as well as “walk” along and across the edges of the modelled road network (cf. [COLOMBO D5.2, 2014]). All other traffic participants – including bicycles – are modelled as a “vehicle” that may be occupied by persons.

The physical attributes of a vehicle, including its length, width, max. velocity, max. acceleration and deceleration ability, and other, are defined within a “vehicle type”. A single “vehicle type” may be shared by an arbitrary number of vehicles. Besides assigning a certain type to a vehicle, SUMO allows to use so-called “vehicle type distributions”. Such a distribution allows to include several “vehicle types”, each with additional information about the probability to be selected<sup>9</sup>. CTAES uses this mechanism to model vehicle populations at different granularity levels and for different years.

Within the scenario descriptions, the modelled modes of transport use the names of carrier modes as listed above. When being executed, an additional file that includes the according “vehicle type” definitions is loaded. Table 5.3 lists the available vehicle type files.

**Table 5.3: Available vehicle type definition files.**

Option Value	Content
plain	A single, “basic” vehicle per carrier type; uses SUMO’s default parameters <sup>10</sup>
2010	Resembles emission population for year 2010 as discussed below
2014	Resembles emission population for year 2014 as discussed below
2020	Resembles emission population for year 2020 as discussed below
2030	Resembles emission population for year 2030 as discussed below
2040	Resembles emission population for year 2040 as discussed below

The given vehicle type files form an additional dimension to the ones of the used scenario sets. It is assumed that only one distribution (plain) is used within the development and initial benchmarking of a traffic light control, the others are afterwards to compute the performance over years. Both aspects – emission behaviour and dynamics – are discussed in the following.

### *Representation of Emission Fleet Composition*

For emission modelling, the vehicle fleet is subdivided into vehicle groups with homogenous emission behaviour. A common method of fleet segmentation is to differentiate by following criteria:

- vehicle category (e.g.: passenger cars, light duty vehicles, rigid trucks, ...)
- engine concept (e.g. gasoline, diesel)
- size class (differentiating factor: capacity or maximum allowed gross weight) and
- emission standard (legislation which was applicable at the vehicles type approval, e.g. “EURO 5”)

A vehicle segment is for example a “rigid truck with diesel engine, gross vehicle weight with more than 18 tons, emission standard EURO 5”.

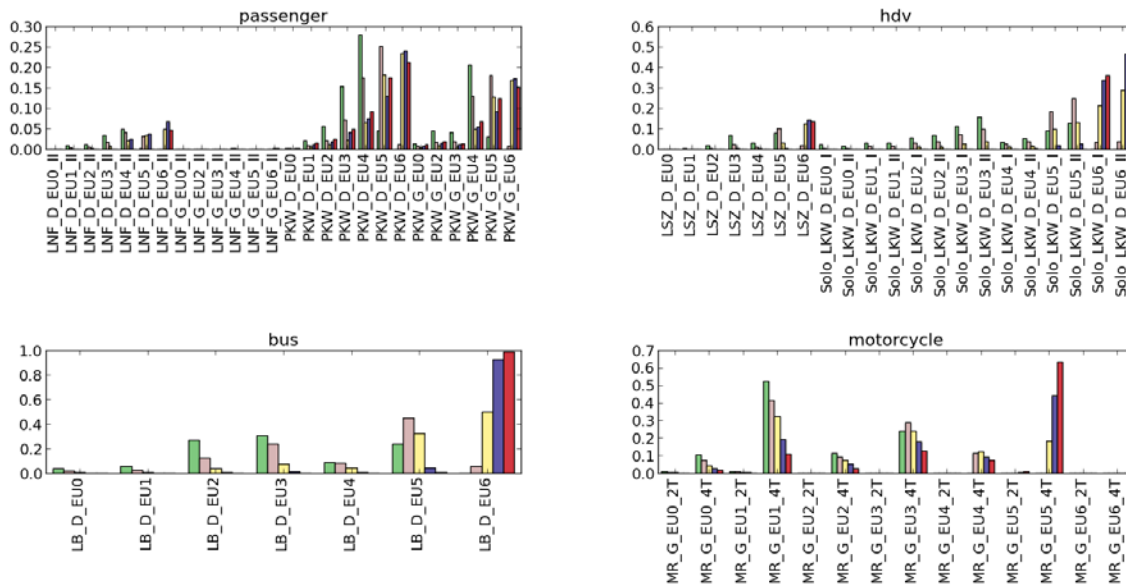
<sup>9</sup> Further information are available at

[http://sumo.dlr.de/wiki/Definition\\_of\\_Vehicles,\\_Vehicle\\_Types,\\_and\\_Routes#Route\\_and\\_vehicle\\_type\\_distributions](http://sumo.dlr.de/wiki/Definition_of_Vehicles,_Vehicle_Types,_and_Routes#Route_and_vehicle_type_distributions)

<sup>10</sup> Given at [http://sumo.dlr.de/wiki/Definition\\_of\\_Vehicles,\\_Vehicle\\_Types,\\_and\\_Routes#Vehicle\\_Types](http://sumo.dlr.de/wiki/Definition_of_Vehicles,_Vehicle_Types,_and_Routes#Vehicle_Types)

For the overall emission output on the street network the shares of the different fleet segments on the overall mileage are relevant. These numbers can be determined by fleet models, which are based on registration statistics and on additional functions, which e.g. depict the mileage dependency on vehicle age and on the road category. Due to the “natural” fleet renewal the fleet composition according to emission relevant fleet segments varies significantly over the years as old vehicles are replaced by new technologies in a permanent process.

The data on fleet composition used in COLOMBO is based on results from the model NEMO (Network Emission Model, see [Dippold et al., 2012] and [Rexeis and Hausberger, 2008], which describes the Austrian fleet until the year 2040. Other distributions that match other European countries could be retrieved from inventory models, such as HBEFA [INFRAS, 2014]. The data from NEMO are used in COLOMBO due to being the best – most exact and least aggregated – data set available in the project. Figure 5.5 shows the shares of vehicle segments as a function of the year.



**Figure 5.5: Probabilities of a vehicle type emission classes in dependence to the year.**

Pedestrians and bicycles are not considered herein, as both groups of traffic participants generate negligible emissions.

### *Representation of Vehicle Dynamics*

SUMO holds defaults for a large variety of vehicles<sup>11</sup>. The definitions “plain”, and “2010”-“2040” reuse these defaults as shown in Table 5.4. The parameters shown in this Table cover pure physical properties of the vehicles as well as parameters of the so-called car-following model that control the simulation behaviour. Explicitly, the parameters have the following meaning:

- vClass: an abstract class used by SUMO to determine which lanes may be used (e.g. it allows to model dedicated bus lanes)
- Sizes: the physical sizes of the participant type instances
- minGap: the average distance to a leader when standing
- $a_{\max}$ : the maximum acceleration
- $d_{\max}$ : the maximum deceleration

<sup>11</sup> See [http://sumo.dlr.de/wiki/Vehicle\\_Type\\_Parameter\\_Defaults](http://sumo.dlr.de/wiki/Vehicle_Type_Parameter_Defaults)

- $v_{max}$ : the maximum velocity

Table 5.4: Vehicle type defaults.

Type	vClass	Sizes (length, width, height <sup>12</sup> )	minGap	$a_{max}$	$d_{max}$	$v_{max}$
pedestrian	pedestrian	0.215 m 0.478 m 1.719 m	0.5 m	1.5 m/s <sup>2</sup>	2 m/s <sup>2</sup>	5.4 km/h
bicycle	bicycle	1.6 m 0.65 m 1.7 m	0.5 m	1.2 m/s <sup>2</sup>	3 m/s <sup>2</sup>	20 km/h
passenger	passenger	4.3 m 1.8 m 1.5 m	2.5 m	2.9 m/s <sup>2</sup>	7.5 m/s <sup>2</sup>	180 km/h
hdv	truck	7.1 m 2.4 m 2.4 m	2.5 m	1.3 m/s <sup>2</sup>	4 m/s <sup>2</sup>	130km/h
bus	bus	12 m 2.5 m 3.4 m	2.5 m	1.2 m/s <sup>2</sup>	4 m/s <sup>2</sup>	85 km/h

## 5.5 PI Computation

SUMO is capable to write a plethora of simulation results. The work on CTAES concentrated on the requirement in section 5.1 by generating the performance indicators (PIs) defined in D1.1 and in chapter 4 of this deliverable. PIs are computed directly after a simulation run by the `ctaes_execute.py` script. The script read some outputs generated by the simulation, aggregates the measures, computes statistics, and writes them into the data base. A large number of other measures than those described in the following could be employed. Those that were found to be interesting for evaluating traffic lights may be enabled when starting an evaluation optionally.

The “tripinfo” output delivers the following per-vehicle information for each vehicle’s summed up trip. `ctaes_execute.py` uses this output to compute global, “network-wide” measures. In addition, this script uses the information about the type of the vehicle to build per-type measures. Among other information, the `tripinfo` output includes:

- **duration:** the duration of the trip/journey (in [s])
- **meanSpeed:** the average velocity (in [m/s])
- **routeLength:** the length of the route (in [m])
- **waitSteps:** the number of simulation steps the vehicle was standing (speed < .1m/s, in [#])
- **CO:** the amount of emitted CO (in [mg])
- **CO<sub>2</sub>:** the amount of emitted CO<sub>2</sub> (in [mg])
- **HC:** the amount of emitted HC (in [mg])
- **PM<sub>x</sub>:** the amount of emitted PM<sub>x</sub> (in [mg])
- **NO<sub>x</sub>:** the amount of emitted NO<sub>x</sub> (in [mg])
- **fuel:** the amount of used fuel (in [ml])

<sup>12</sup> The height is currently not used.

The following aggregation step computes the following statistical derivatives for the whole network:

- **avg:** the average
- **stddev:** the standard deviation
- **median:** median
- **q25:** the 25% quantile
- **q75:** the 75% quantile
- **sum:** the sum

In addition, different vehicle types can be distinguished, allowing to compute the aforementioned PIs and their derivatives individually for each vehicle class according to the requirement in section 5.1. This is done by default by the execution script. The output covers almost all PIs defined in section 4.1 and [COLOMBO D1.1, 2014], beside the User Acceptance, as shown in Table 5.5.

**Table 5.5: Covered Road Network PIs as defined in D1.1**

Criteria / PI	Definition / Comments
<b>Objective: Mobility</b>	
total travel time	sum:travelTime
mean travel time	avg:travelTime
mean speed	avg:meanSpeed
total waiting time	sum:waitingTime
mean waiting time	avg:waitingTime
mean number of stops	N/A
total distance travelled	sum:routeLength
mean distance travelled	avg:routeLength
<b>Objective: Resource Efficiency</b>	
mean fuel consumption	avg:fuel
network saturation (I/C-ratio)	N/A
<b>Objective: Environmental Impact</b>	
mean exhaust emissions for pollutant x	avg:x
mean noise emissions	N/A

At the time being, no user-defined aggregation intervals are supported for network-wide measures. A future modification could allow this by using a the “emissions” simulation output that lists all vehicles in all time steps together with a large variety of their instantaneous states.

Besides these measures that cover the complete simulation area, further intersection-related PIs can be gathered for each of the scenario sets’ traffic lights. These PIs are collected by observing the areas in front of the traffic lights. Here, two aggregation types are used: a) per cycle, and b) per user-given aggregation interval.

The system is as well capable to compute the single performance indicator described in Chapter 4. This is done by calling the original PI computation tool. The data the PI computer uses are generated during the respective simulation. After its completion, the results written by this PI computation tool are then read from the file and added to the database. The inclusion of the PI

computation script is has to be enabled by the user by adding the `--compute-singlePI` switch when starting `ctaes_execute.py`.

## 5.6 Implementation and Deployment

The system was implemented in the well-established and mature programming language Python. This assures a high portability as Python itself is available for most modern operating systems. In the following, some aspects of the software will be discussed.

### Availability and Licensing

Being stored in the project's source code control, the evaluation system is currently available for COLOMBO partners only. Licensing and availability is not yet defined. Both will be targeted within oncoming Dissemination and Exploitation activities.

### System Requirements

As stated, the system's portability is assured by using the Python programming language.

### Scenario Sets Implementation

Technically, each scenario is represented as an own Python class that mainly defines the files it consists of as well as other scenario parameters such as the time the scenario needs to be filled with vehicles and its end time. Scenario sets are represented as Python classes as well, albeit rather containing program code that iterates over the respectively examined traffic characteristic (see chapter 3) or allowing to retrieve meta-information about the iteration process than referring to files. The code dependency between scenario sets and scenarios as well as the attributes of these classes are depicted in Figure 5.6.

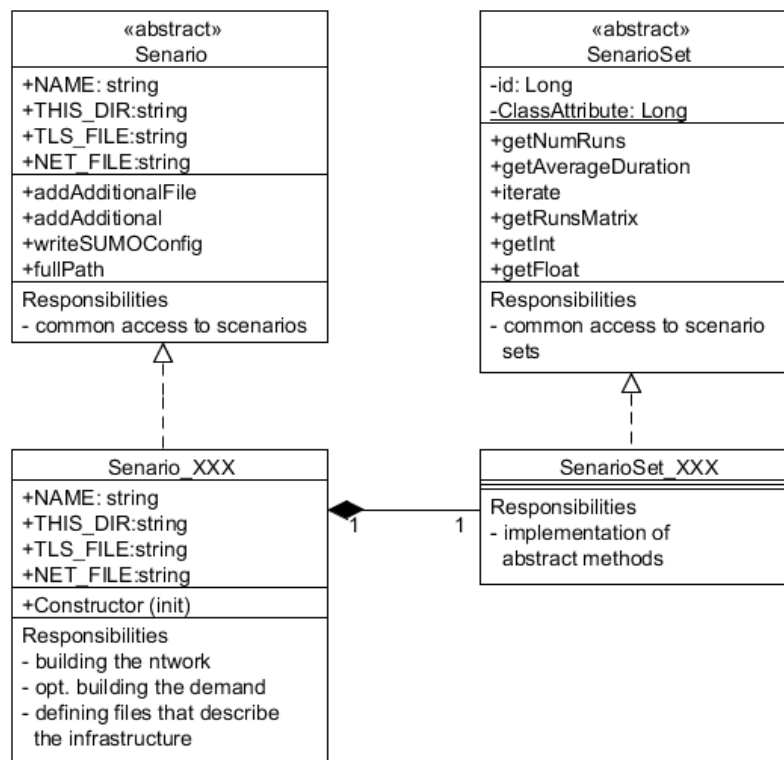


Figure 5.6: Class diagram for scenarios and scenario sets.



## User Documentation

No user documentation besides this document is available, yet.

## Applicability for other Simulation Packages

One requirement to the TLS Evaluation System was to make it usable in conjunction with other traffic simulations than SUMO. As described, the CTAES prepares the scenario to execute first, then runs the used simulation, and finally reads the results to compute performance indicators used internally. Both parts are implemented in a modular way. To use a different traffic simulation than SUMO, the following steps are necessary:

- implement data export modules that translate the scenario descriptions into input files for the used simulation;
- implement data import modules which supply measures for the used performance indicators;
- implement a module which starts the used simulation.

## Data Handling

The system stores and retrieves results in a SQLite<sup>13</sup> database. As SQLite is included in Python, no additional installation is required. SQLite uses files to store data what offers a range of benefits. The major one is the possibility to release pre-computed performance indicators for different traffic light control algorithms, such as a plain static phase plan (“fixed-cycle”) or detector-based vehicle-actuated controls. In addition, the user may have different database copies for different versions of her/his algorithm. SQLite is supported by different applications that read and/or to modify its database files.

CTAES database schema consists of three tables only. The first, named “runs” contains the meta data of the performed simulation runs. The entries in this table consist of three fields:

- **id:** the simulation run’s unique identifier (ID)
- **key:** the name of a parameter of the simulation run
- **value:** the value of this parameter

Each simulation run is defined using a unique ID. The given key/value pairs allow to assign it to one of the executed scenarios. E.g. a simulation run for a certain flow combination (400 vehicles/h in north-south, 1200 vehicles/h in east-west direction) from the scenario set “iterateFlowsNA” (see Appendix C) for the “static” algorithm (fixed time control) would be defined using the entries given in Table 5.6.

**Table 5.6: Example for entries in the “run” table that define a certain simulation run (see text).**

Database Entry			Description (not included in the database)
id	key	value	
0	scenario	BasicCross	Defines the used scenario
0	f1	400	Defines the north-south flow
0	f2	1200	Defines the east-west flow
0	tls_algorithm	static	Defines the used traffic light algorithm

<sup>13</sup> From <http://www.sqlite.org/> (22.08.2014): SQLite is a software library that implements a self-contained, serverless, zero-configuration, transactional SQL database engine.

It is possible to have more than one simulation run of same type being stored in the database to ensure the statistical validity.

The second table contains the results from each simulation run, stored using the following fields:

- **id:** the simulation run's unique identifier (ID)
- **denominator:** the name of the artefact the measure belongs to (see below)
- **key:** the name of the PI (composed of an abbreviation of the statistical property (mean, median, average, etc.) and the measure's name, see also section 5.5)
- **value:** the value of the performance indicator

Each simulation run contains one entry per performance indicator (and statistical property), denominator, and interval. The "denominator" is a text label that distinguishes either vehicles classes or network elements. Per default, measures are generated for all vehicle classes (denominator="global") and for every vehicle class individually (the denominator matches the vehicle class as given in section 5.4).

The third table is an enlarged version of the second one, extended by the begin time and the end time of the period:

- **id:** the simulation run's unique identifier (ID)
- **denominator:** the name of the artefact the measure belongs to
- **aib:** the begin of the (aggregation) interval
- **aie:** the end of the (aggregation) interval
- **key:** the name of the PI
- **value:** the value of the performance indicator

## 5.7 First Insights

The system was so far used to compare some implemented traffic light algorithms. It was found to be very nice to use, albeit being run from the command line. The resulting views on performance in dependency to certain traffic characteristics are often easy to understand and confirm expectations taken a-priori.

From the user perspective, the execution speed is important – having the results in one hour is nice. Waiting for a complete scenario set for one week to finish is only acceptable if the results are very reliable and meaningful. The bigger scenario sets that were developed (see Appendix C) should be therefore revisited for inspecting whether they really cover the stated traffic characteristic. Of course, every simulation system is wanted to be faster in execution time. A lot of time is used to parse the outputs generated by SUMO, computing the PIs, and writing the latter to the database. It may be interesting to optimize the database calls. As well, the bigger scenarios could be decomposed or sets consisting of the most important scenarios of a scenario set could be set up.

It should be mentioned that a lot of measures are collected, so that the results databases can get relatively big in a relative short time. Albeit a post-processing is performed that removes the verbose XML notation and joins individual values into summarizing aggregates, the databases may get easily some ten MB in size what slows the access times. This is especially true when time aggregations and per-lane measures are collected. When compared to the raw data generated by the simulation that can easily be some GB in size, this is still a big reduction. Nonetheless, it should be verified whether really all PIs with all of their statistical properties must always be kept.

The given scenario sets test a traffic light algorithm against a broad variety of problems. Though, some of them should be revisited and/or verified. Especially the determination whether a synchronisation takes place or not is complicated to cover, because it is ambiguous to a plain adaptation to the incoming vehicle streams. This has to be investigated in the future.



## 6 Summary

The task to develop a traffic light evaluation and benchmarking approach was covered by two major results: a) a new single performance indicator was presented that combines measures describing traffic efficiency, user perception, and environmental issues into a single value and b) a software application that uses a set of well-chosen simulation scenarios to evaluate how good a traffic light performs under certain situations. Both parts are covered by determining nowadays methods for benchmarking traffic lights and by evaluations of real-world traffic data to obtain a deeper insight about the traffic characteristics that had to be modelled.

The newly developed performance indicator allows to quantify a traffic light's performance using a single measure only by joining and weighting a broad set of performance indicators from different domains. Besides being applicable in the context discussed herein – the development and benchmarking of traffic light algorithms – it is as well applicable for benchmarking real-world traffic lights. As well, it allows optimizing a traffic light using algorithms that rely on obtaining a single measure only.

The traffic light algorithm evaluation system is based on a large variety of so-called scenario sets that perform iterative simulations, testing the behaviour of the algorithm under different infrastructure and traffic flow settings. We assume this system to be well beyond nowadays state-of-the-art where arbitrary and often incompletely defined scenarios are used. Given the scenarios sets presented here, a developer is capable to investigate which characteristic influences the performance to which degree.

So far, the included scenario sets cover several traffic characteristics and are assumed to be well-chosen and well-designed. They can be described using few parameters (as done during their presentation) and most of them pose the TLS against easy ("trivial") as well as against complicated problems. Still, traffic characteristics as well as the implicitly assumed features of a traffic light – adaptation to demands, their changes over longer periods as well as small fluctuations, to amounts of turning vehicles, to special vehicle types, synchronisation, etc. – are covered at different depths. In the case one of those gets into the focus – the evaluated traffic light control targets especially a certain influence, such as keeping the outflow lanes free – further scenario sets should be added. Such added scenarios can be applied to basic, established control methods, such as a fixed-cycle traffic light or an actuated control, out-of-the-box. New scenarios should be included into the standard scenario base after being verified. The already given scenario sets should be revisited as well to determine whether they really cover what they are assumed to.

The presented system is new and was not used besides COLOMBO scopes, yet. A first version is assumed to be released in the next months. It is thereby still a scientific tool; review steps should be performed and feedback for enhancements is needed. Further, probably major, changes based on user interaction are assumed (and hoped) to be performed in the future. Such interaction with the traffic community has begun and will be continued.

Using a common interface to synthetic and real-world scenarios is a feature that may be reused at other places than evaluation of traffic lights. It already has been used for the development and evaluation of other COLOMBO solutions the local emissions monitoring system (presented in [COLOMBO D1.2, 2014]) and the V2X-based traffic surveillance (D1.2 as well). As well, it was used to obtain scenarios used during the off-line configuration of traffic lights (see D2.4 "Performance of the Traffic Light Control System for different Penetration Rates" and D3.2 "Results of the offline Configuration and Tuning of the emergent Behaviour"). Several further applications are possible, the major to name are automated tests of the used traffic simulation itself.

## Appendix A – References

- [Akcelik, 2000] Akcelik, Rahmi: Traffic Management Performance – Development of a Traffic Frustration Index, Section 3, Publication AP-R160, Austroads, Sydney, 2000, ISBN 0 85588 552 1.
- [Axhausen, 2014] Axhausen, K.: Comparison of estimates of travel time losses on high capacity roads, TRB Annual Meeting, 2014, Washington D.C.
- [Bernard and Axhausen, 2010] Bernard, M. and K.W. Axhausen: Ein neuer Ansatz für standardisierte Ganglinien, Arbeitsberichte Verkehrs- und Raumplanung, 540, IVT, ETH Zürich, Zürich.
- [Bijl et al., 2011] Bijl, B. van der, Vreeswijk, J. D., Bie, J. and Berkum, E. C. van; Car drivers' perception and acceptance of waiting time at signalized intersections. 14th IEEE ITSC Conference, Washington DC, 5-7 October 2011.
- [BMVBS, 2012] Bundesministerium für Verkehr, Bau und Stadtentwicklung (publisher): Technische Lieferbedingungen für Streckenstationen (TLS), Berlin, 2012. [http://www.bast.de/DE/FB-V/Publikationen/Regelwerke/Unterseiten/V5-tls-2012.pdf?\\_\\_blob=publicationFile&v=1](http://www.bast.de/DE/FB-V/Publikationen/Regelwerke/Unterseiten/V5-tls-2012.pdf?__blob=publicationFile&v=1), last accessed: 09.11.2014
- [BMVBW, 2003] Bundesministerium für Verkehr, Bau- und Wohnungswesen (publisher): Bundesverkehrswegeplan 2003, Grundlagen für die Zukunft der Mobilität in Deutschland. Beschluss der Bundesregierung vom 2. Juli 2003, Berlin, 2003.
- [BMVIT, 2014] Bundesministerium für Verkehr, Innovation und Technologie: Straßenverkehrsordnung Österreich, BGBl.Nr. 159/1960 zuletzt geändert durch BGBl. I Nr. 27/2014, Wien.
- [COLOMBO D1.1, 2014] COLOMBO project consortium: Scenario Specifications and Required Modifications to Simulation Tools, deliverable 1.1, February 2014.
- [COLOMBO D1.2, 2014] COLOMBO project consortium: Data Collection and Dissemination for Low Penetration Systems, deliverable 1.2, November 2014.
- [COLOMBO D5.1, 2014] COLOMBO project consortium: Prototype of overall System Architecture and Definition of Interfaces, deliverable 5.1, 2014.
- [COLOMBO D5.2, 2014] COLOMBO project consortium: Traffic Simulation Extensions, deliverable 5.2, May 2014.
- [Dippold et al., 2012] Dippold M., Rexeis M., and Hausberger S.: NEMO – A Universal and Flexible Model for Assessment of Emissions on Road Networks. 19th International Conference “Transport and Air Pollution“, 26. – 27.11.2012, Thessaloniki.
- [FGSV, 1997] Forschungsgesellschaft für Straßen- und Verkehrswesen FGSV (publisher): Empfehlungen für Wirtschaftlichkeitsuntersuchungen von Straßen (EWS), FGSV Verlag GmbH, Köln, 1997.
- [FGSV, 2009] Forschungsgesellschaft für Straßen- und Verkehrswesen FGSV (publisher): HBS – Handbuch für die Bemessung von Straßenverkehrsanlagen, as amended by 2009, FGSV Verlag Köln, ISBN 3937356444, 2001/2009.
- [FGSV, 2010] Forschungsgesellschaft für Straßen- und Verkehrswesen FGSV (publisher): RiLSA: Richtlinien für Lichtsignalanlagen – Lichtzeichenanlagen für den Straßenverkehr, 2010.
- [FGSV, 2015] Forschungsgesellschaft für Straßen- und Verkehrswesen FGSV (publisher): Richtlinien für Wirtschaftlichkeitsuntersuchungen an Straßen (RWS), expected in 2015.

- [FHWA, 2008] U.S. Department of Transportation, Federal Highway Administration (publisher): Surrogate Safety Assessment Model (SSAM). Techbrief, FHWA-HRT-08-049, 05/2008, <http://www.fhwa.dot.gov/publications/research/safety/08049/>, last accessed 09.11.2014.
- [Fietsberaad 2004] Fietsberaad: De fietsvriendelijkheid van verkeersregelininstallaties publication 4, Utrecht, 2004. [http://www.fietsberaad.nl/library/repository/bestanden/nPublicatie\\_41.pdf](http://www.fietsberaad.nl/library/repository/bestanden/nPublicatie_41.pdf)
- [FSV, 2013] Österreichische Forschungsgesellschaft Straße – Schiene – Verkehr: Evaluierung von Verkehrslichtsignalanlagen, RVS 05.04.35, Wien, 2013.
- [Hunter, 2010] Hunter, B.: Assessment of the traffic quality at signalized intersections for all travelers, Master Thesis, TU Darmstadt, 2010.
- [INFRAS, 2014] INFRAS, Handbook Emission Factors for Road Transport 3.2, Bern, available online at <http://www.hbefa.net>
- [Kittleson and Roess, 2001] Kittleson, W., and R. P. Roess: Highway Capacity Analysis after Highway Capacity Manual 11 2000. Transportation Research Record, Vol. 1776, 2001, pp. 10-16.
- [Lämmer and Däumler, 2014] Lämmer, S., and Däumler, M.: Ansätze zur statistischen Modellierung von Nachfrageschwankungen auf Stunden- und Minutenskalen im Stadtstraßenverkehr, HEUREKA '14, Stuttgart, 2014.
- [Martin, 2006] Martin, A., Factors influencing pedestrian safety: a literature review, TRL Limited, 2006.
- [Müllner, 2013] Müllner, D.: fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python, Journal of Statistical Software 53 (2013), no. 9, 1–18, <http://www.jstatsoft.org/v53/i09/>, last accessed: 09.11.2014
- [Niebel, 2013] Niebel, W.: Cost-Benefit-Based Implementation Strategy for Green Light Optimised Speed Advisory (GLOSA), In: Communications in Computer and Information Science, 395, pp. 313-320. Springer-Verlag Berlin Heidelberg, 2013. ISBN 978-3-642-41646-0. ISSN 1865-0929.
- [Pinkofsky, 2006] Pinkofsky, L. (2006) Typisierung von Ganglinien der Verkehrsstärke und ihre Eignung zur Modellierung der Verkehrsnachfrage, Ph.D. Thesis, Technical University Braunschweig, Braunschweig.
- [PRESTO, 2009] PRESTO project consortium: Give Cycling a Push: Implementation Fact Sheet, project publication, 2009-2012.
- [Pulugurtha and Kusam, 2011] Pulugurtha, S.S. and P.R. Kusam: An Assessment of Multimodal Level-of-Service as a Performance Measure for Signalized Intersections, TRB Annual Meeting 2011, Washington D.C.
- [Rexeis and Hausberger, 2008] Rexeis, M., Hausberger, S., Trend of vehicle emission levels until 2020 – Prognosis based on current vehicle measurements and future emission legislation, Atmospheric Environment (2008), doi:10.1016/j.atmosenv.2008.09.034
- [Schiller et al., 2011] Schiller, C., Zimmermann, F., Bohle, Wolfgang: Hochrechnungsmodell von Stichprobenzählungen für den Radverkehr, Schlussbericht, FE 77.495/2008, TU Dresden.
- [Smart et al., 2014] Smart, M., McCann, H., and M. Brozen: What's a Passing Grade? A Synthesis and Review of 4 Multimodal Street Performance Measures, TRB Annual Meeting 2014, Washington D.C.
- [TRB, 2010] Transportation Research Board: Highway Capacity Manual HCM2010, ISBN 978-0-309-16077-3, Washington D.C.
- [VSS, 1999] VSS: Leistungsfähigkeit, Verkehrsqualität, Belastbarkeit; Grundlagennorm, Norm, SN640 017a, Swiss Association of Road and Transport Professionals (VSS), Zurich.



- [VSS, 2008] VSS: Leistungsfähigkeit, Verkehrsqualität, Belastbarkeit; Knoten mit Lichtsignalanlagen, Norm, SN640 023a, Swiss Association of Road and Transport Professionals (VSS), Zurich.
- [VSS, 2010] VSS (2010) Ganglinientypen und durchschnittlicher werktäglicher Verkehr, Norm, SN 640 005b, Swiss Association of Road and Transport Professionals (VSS), Zurich.
- [Wietholt, 2009] Wietholt, T.: Einsatzbereiche Grüner Wellen und verkehrsabhängiger Steuerungen. PhD Thesis, ISSN 1437-8299, Schriftenreihe des Lehrstuhls für Verkehrswesen Ruhr-Universität Bochum, Vol. 33, Bochum 2009.
- [Yang, 2012] Yang, X., et. al., Crossing at a red light: Behavior of cyclists at Urban Intersections, Discrete Dynamics in Nature and Society, Volume 20122, Article ID 490810.

## Appendix B – Load Curves Clustering Dendrograms

The following dendrograms show how the time lines of flows across inductive loops that were used in section 3.1 differ (per observation site).

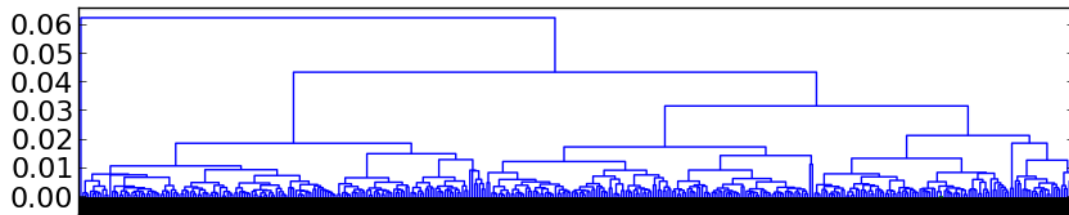


Figure B.1: Dendrogram of Bologna time lines similarity.

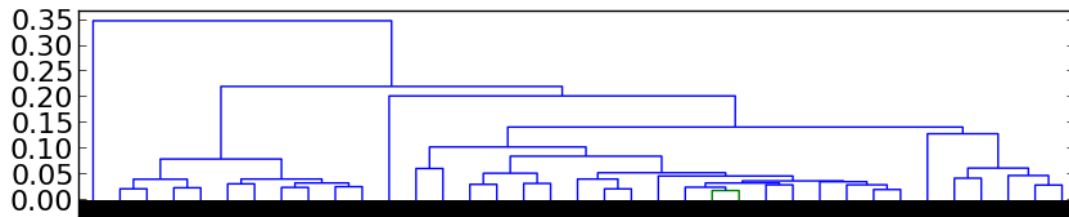


Figure B.2: Dendrogram of Cologne time lines similarity.

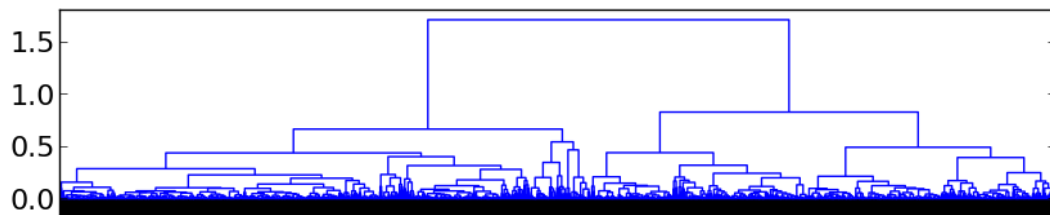


Figure B.3: Dendrogram of Brunswick time lines similarity.

## Appendix C – Scenario Sets

In the following, the scenario sets are discussed individually. This is done by comparing two different traffic light algorithms:

- a fixed-time control;
- an actuated control.

The fixed time control switches the traffic lights based on a pre-given program with constant duration. The durations themselves will be given in description of the respective scenario set. The actuated traffic light algorithm starts with the same signal plan. In every phase, the duration of the green light is increased as long as a vehicle is sensed by the inductive loops not longer than a given time threshold (the default of 3.1 s is used in the subsequent). The length of a green duration is bound by a maximum time. The inductive loops are placed automatically by the algorithm's implementation using the default distance of 3 m in the following.

Both algorithms are common in the real world. Still, one should mention that the actuated traffic light algorithm used herein does not preserve a given cycle time.

The following presentation is not meant to value the given program plan. Neither the absolute values of the given measures are of a bigger interest. Rather, it is attempted to show whether and how a traffic light algorithm reacts on the change of the respectively addressed traffic characteristic. The scenario sets have been chosen to span over reasonable and relevant ranges of the addressed characteristic.

Every subsequent presentation of a scenario set includes a figure that shows three matrices. The left and the middle ones show the performance of the algorithm, given the according scenario. The right one shows the difference between both, correctly:

$$dt_{\text{waiting}} = t_{\text{waiting,actuated}} - t_{\text{waiting,static}} \quad (\text{C-1})$$

where

$t_{\text{waiting,actuated}}$ : the waiting time when using the actuated control

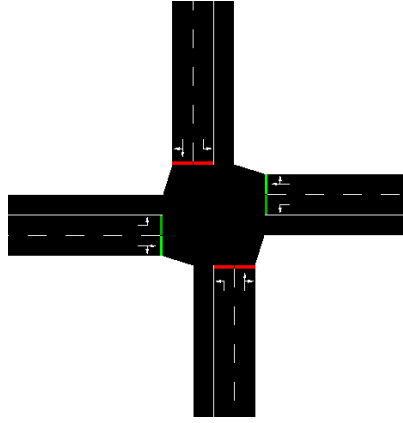
$t_{\text{waiting,static}}$ : the waiting time when using the fixed-time control

Please note that only the performance regarding the waiting time will be shown in the following for reducing the size of this document. This measurement was used as it is assumed to show the development of the performance within a scenarios set best. Other evaluation figures are only shown if they are needed to explain the respective scenario set. If not stated differently, all figures use average values from 12 simulation runs.

### *A: Iterations over static Flows with no a-priori Adaptation (“iterateFlowsNA”)*

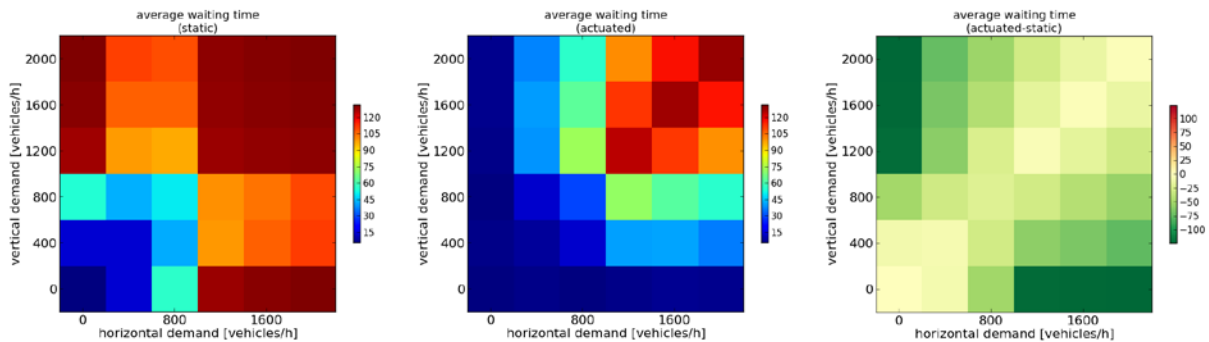
Traffic lights should adapt the green times to a changing flow. One of the most trivial methods is to confront the traffic light with different amounts of flows that are fixed for the duration of a simulation run and observing whether the traffic light is capable to adapt itself to these static demands.

This scenario set realises such investigations using a simple single intersection. All scenarios start with the same green light durations for both directions: the cycle length is 80 s, and each direction gets green for 32 s. All vehicles run straight. The intersection layout is depicted in Figure C.1.



**Figure C.1: The “Basic Cross” intersection.**

The evaluated traffic light control is not only responsible to react on small traffic flow changes (“fluctuations”), but is itself responsible for assigning the available green time to the given, static demand. Figure C.2 shows the behaviour of a fixed-cycle traffic light in comparison to a default actuated traffic light algorithm.



**Figure C.2: A comparison of the performance of a fixed-cycle and an actuated traffic light for the “iterateFlowsNA” scenario set.**

Figure C.2 shows that the lack of adaptation of a fixed-cycle TLS yields in an increase in average waiting times as soon as one of the streams is higher than about 1000 vehicles / hour. The actuated traffic light reacts to the demand changes and a bigger increase in waiting times occurs only if both directions have a flow above about 1000 vehicles / hour.

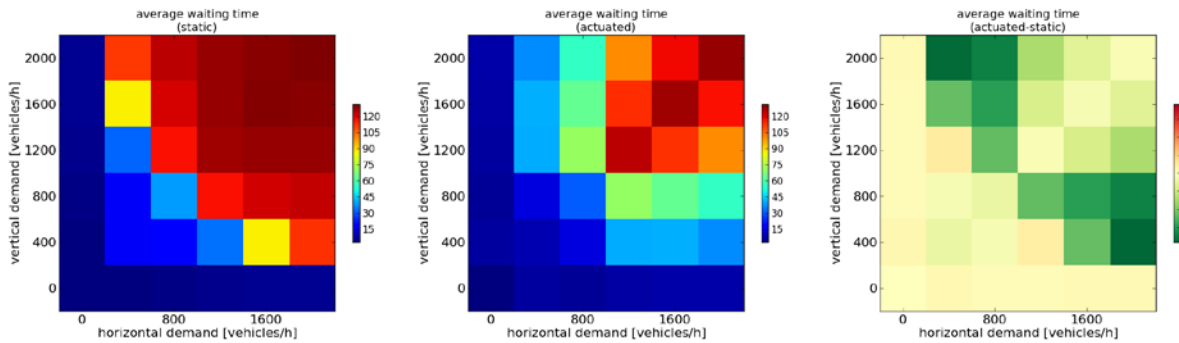
Notes about the scenario set:

- A fixed-plan traffic light should be worse in all cases (given non-constant time headways between vehicles) despite the diagonal when compared against another TLS algorithm, because in all other cases the traffic light control has to adapt the green times to different amounts of vehicles running in different directions.
- Not all traffic light control algorithms may be capable to adapt the green times in such a wide range, simply because they may be designed to cope with smaller bands.

### ***B: Iterations over static Flows with a-priori Adaptation (“iterateFlowsA”)***

No real-world traffic light system is assumed to adapt itself to such a different magnitude of traffic flows as evaluated by the previously mentioned “iterateFlowsNA” scenario set. Instead, weekly time plans (see also the following scenario set “RiLSAllLoadCurves”, e.g.) are used that compensate big flows changes. The “iterateFlowsA” scenario set determines how well a traffic light control is capable to adjust itself to small changes and/or fluctuations. It is based on the “iterateFlowsNA” scenario set, but for each simulation run (scenario), the green times (per

direction) are computed by dividing the available green time into proportions that match the ones between the magnitudes of foe streams. This means that the traffic light control starts with green times that are almost perfect (despite modelling aspects and small fluctuations in the demand) for the given demand. The traffic light control is therefore responsible to adapt to such small changes only.



**Figure C.3: A comparison of the performance of a fixed-time and an actuated traffic light for the “iterateFlowsA” scenario set.**

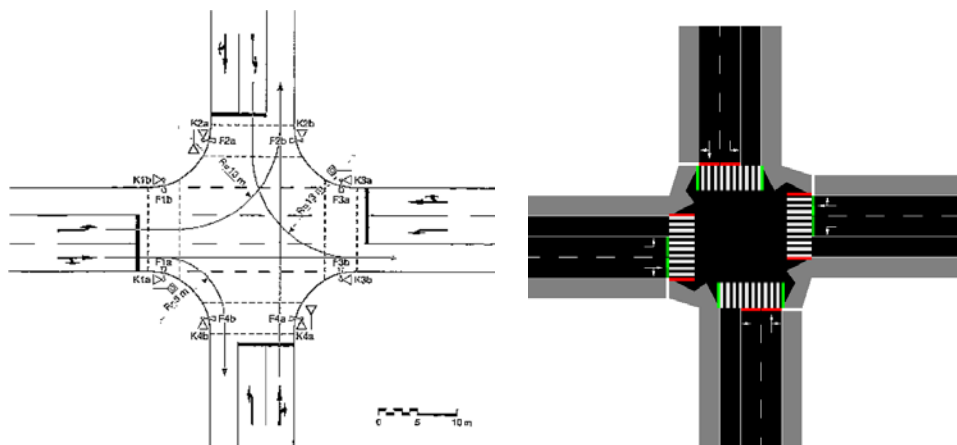
Figure C.3 shows that the adaptation of green times to flows yields in a performance increase for the fixed-cycle TLS when compared to the earlier presented “iterateFlowsNA” scenario set; the threshold of about 1000 vehicle / hour / direction is no longer existing. Still, the traffic light is not performing as well as an actuated one. The actuated traffic light algorithm performs very similar to the “iterateFlowsNA” case.

Notes about the scenario set:

- This scenario set should be applicable and meaningful for all traffic light algorithms.

### *C: Iterations over Daily Load Curves (“RiLSALoadCurves”)*

As discussed in section 3.1, traffic flow usually changes over a day. To obtain realistic scenarios which represent this traffic characteristic, a single-intersection network based on the first example from the examples appendix of the German “RiLSA” handbook is used. Figure C.4 shows this intersection.

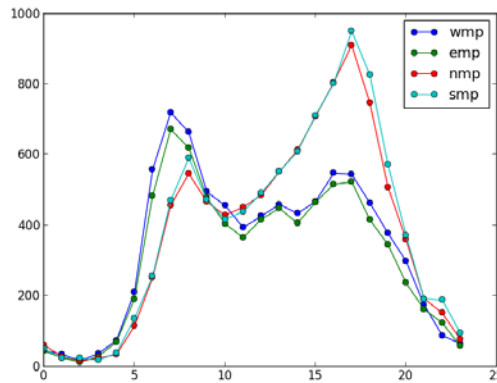


**Figure C.4: First of the RiLSA examples; a) as shown within the RiLSA appendix, b) as simulated.**

The examples in the RiLSA handbook show how traffic light phases that match the regulations are computed. For this purpose, besides the intersection layout, a certain demand is given. The initial idea was to use the given demand, use it for the peak hour, and scale the demands of the remaining hours of a day according to the daily load curves from the “RWS” standard (see section 3.1). Based

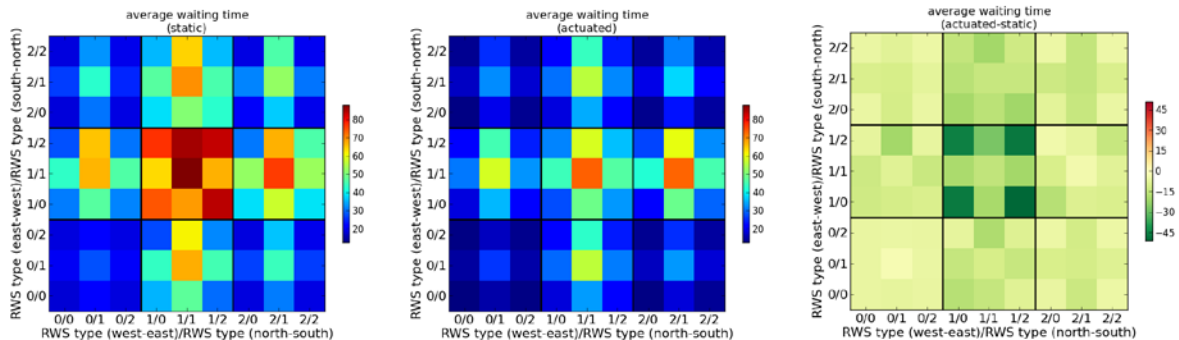
on this, all possible combinations of the RWS curves should be tested within “RiLSA1LoadCurves”.

It turned out, that due to the big differences of given RiLSA flows, the very dominating North/South streams stay above the East/West-streams in all load curve combinations. But the scenario set was built to validate a traffic light algorithm’s response to changes in the dominant direction over the day. Therefore, the given RiLSA flows were scaled before applying the load curves. Figure C.5 shows an example combination of the demands computed this way. The first letter denotes where the respective flow comes from (North, South, West, and East).



**Figure C.5: Resulting flows for combining the curves.**

To cope with demands that change over a day, real-world fixed-cycle traffic lights usually change their programs following a weekly (or daily) switch plan. Within the evaluation suite, only one daily switch plan is realized, as the scenario covers one day only. Still, this plan is computed for every combination of the demands in steps of one hour. As no standardized way to compute the switch plans could be found, an own algorithm was implemented, based on examples from past projects (mainly “ORINOKO”). The algorithm is not presented here.



**Figure C.6: A comparison of the performance of a fixed-time and an actuated traffic light for the “RiLSA1LoadCurves” scenario set.**

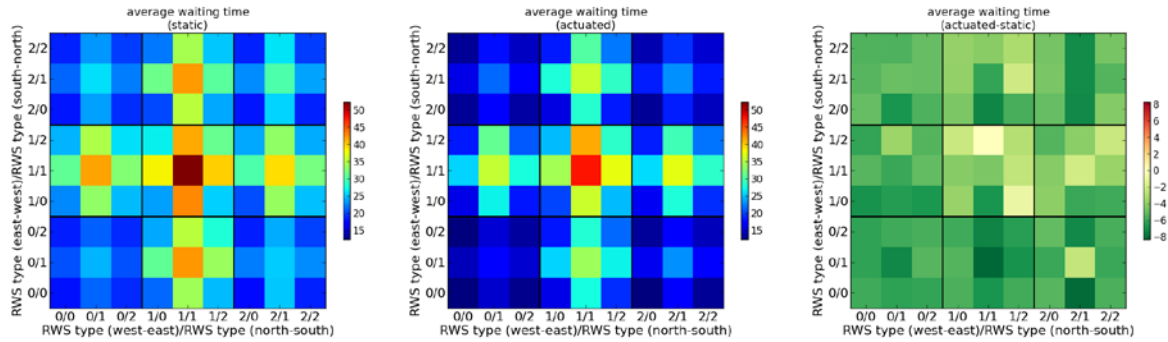
Figure C.6 shows that the occurrence of flows scaled with the second RWS load curve (afternoon peak) yields in biggest waiting times. This behaviour was expected, as in this case the flows are scaled above their initially given values for the morning peak.

Notes about the scenario set:

- Not all traffic light control algorithms may be capable to adapt the green times in such a wide range, simply because they may be designed to cope with smaller bands.
- It may happen that an evaluated TLS algorithm with that does not use a daily switch plan performs worse than the fixed-cycle one that uses them.

### D: Iterations over Sampled Daily Load Curves (“RiLSA1LoadCurvesSampled”)

This scenario set was built to reduce the execution time of the “RiLSA1LoadCurves” scenario set. To achieve this, the original demand is resampled. Every 24<sup>th</sup> second is taken from the original demand to obtain a load curve reduced to 1 h simulation time. As seen in Figure C.7, this yields in a similar performance behaviour as for the “RiLSA1LoadCurves” scenario set presented before. Due to the smaller durations of higher flows the waiting times are compensated more easily, yielding in a lower magnitude of waiting times.



**Figure C.7: A comparison of the performance of a fixed-time and an actuated traffic light for the “RiLSA1LoadCurvesSampled” scenario set.**

Although the obtained patterns seem to be similar to the ones from the original “RiLSA1LoadCurves” scenario set, the patterns of the comparisons between respective fixed-cycle and actuated performance differ significantly. It is thereby questionable that both scenario sets address the same characteristic.

Notes about the scenario set:

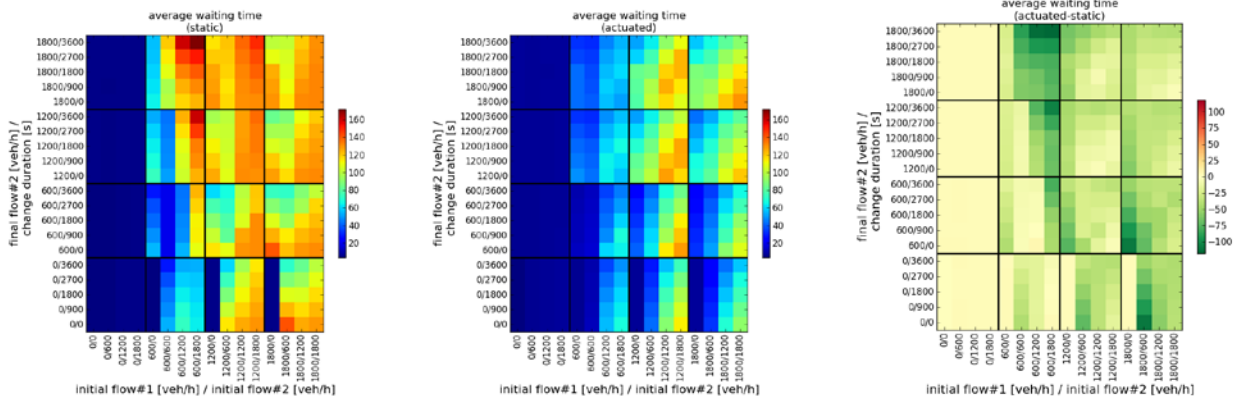
- At the time being, this scenario sets’ applicability is assumed to be the same as for “RiLSA1LoadCurves”.

### E: Singular Demand Changes (DemandStep)

The “DemandStep” is an approach to determine whether and how fast an algorithm adapts to a change in one of the traffic flows. It performs a four-fold iteration:

- level1: static counter flow amount (from 0 vehicles/hour to 1800 vehicles/hour in steps of 600 vehicles/hour)
- level2: initial affected flow’s amount (from 0 vehicles/hour to 1800 vehicles/hour in steps of 600 vehicles/hour)
- level3: final l affected flow’s amount (from 0 vehicles/hour to 1800 vehicles/hour in steps of 600 vehicles/hour)
- level4: duration of the flow’s magnitude change (from 0 s to 3600 s in steps of 900 s)





**Figure C.8: A comparison of the performance of a fixed-time and an actuated traffic light for the “DemandStep” scenario set.**

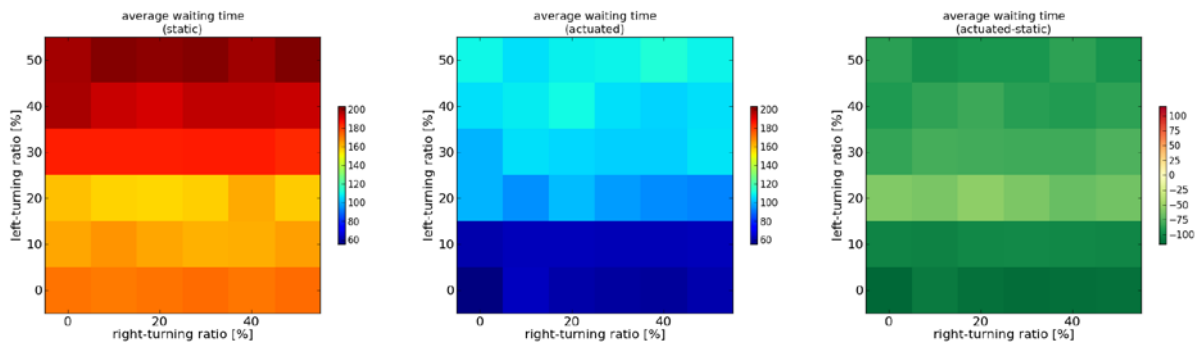
The fixed-cycle traffic light fails in adapting to big traffic flow changes and performs worse with an increase in transition durations. Contrary, the actuated traffic light logic improves in performance if the transition time is increased.

Notes about the scenario set:

- On the one hand, the scenario allows to investigate adaptations to flow changes in a very detailed manner. On the other hand but, a large number of simulation runs is performed making the scenario unhandy to use. One possible solution may be to employ a more sophisticated kind of sampling the iteration space.

#### **F: Turning Ratio Changes (TurnIteration)**

Turning ratios may have a strong effect on traffic lights performance and regarding them may improve a traffic light’s performance. “TurnIteration” evaluates a TLC behaviour in dependence to turning ratios by iterating the amounts of right- and left-turning vehicles between 0% and 50 % in steps of 10 % for the east-to-west stream. The opposite direction is kept at running straight to 100 % to generate a well-defined foe stream. As well, 100 % of the north/south and south/north streams are running straight. The network from “iterateFlowsNA” is used, flows in all directions are kept at 800 vehicle/hour. The results for a fixed-cycle and an actuated traffic light are given in Figure C.9.



**Figure C.9: A comparison of the performance of a fixed-time and an actuated traffic light for the “TurnIteration” scenario set.**

At a threshold between 20 % and 30 % of vehicles turning left, one can see a strong increase in waiting times. The amount of vehicles turning right does not seem to have any influence on the performance – albeit it should be stated that a) no crossing pedestrians are given and b) SUMO does not model deceleration in curves. The actuated traffic light is not capable to compensate an

increased amount of vehicles turning left completely. Still, the development of the waiting times is rather smooth.

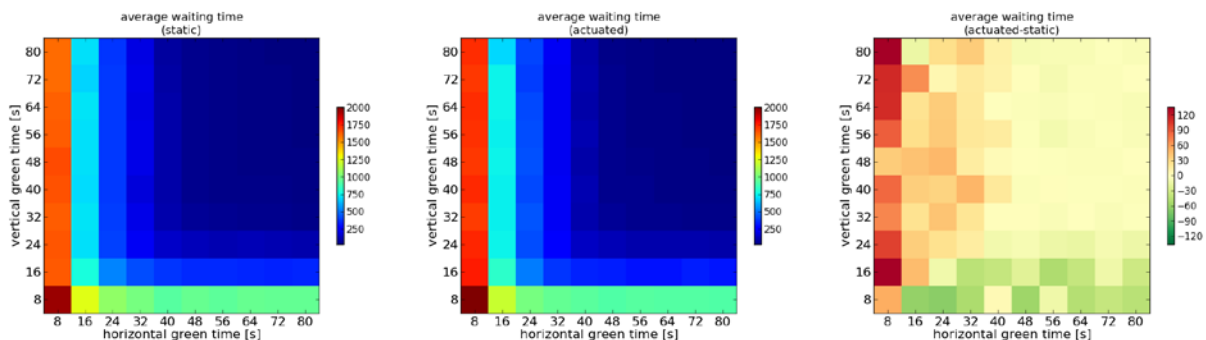
Notes about the scenario set:

- In the real world, traffic light settings do not compensate bigger values of turners; instead, the complete intersection and traffic lights plan would have been adapted when the amount of turning vehicles would become bigger than a certain threshold. This scenario set is therefore very artificial.
- The results should be stated with care unless SUMO does not restraint the velocity of vehicles in curves.
- One could compare this scenario's complexity with the one of "iterateFlowsNA" – a flow is increased. If solving issues of turning vehicles gets more important, further scenario sets that evaluate this characteristic should be added.

### G: Outflow Limitation (RiLSA1Outflow)

Some traffic light algorithms – e.g. the self-organising traffic lights developed in COLOMBO's WP2 – take into regard the situation on roads that start at the intersection (outgoing roads). The motivation is twofold: first, green time that would be wasted for vehicles that may not cross the intersection is given back to streams that may leave the intersections, second, the intersection is kept free as some drivers would enter it even if they cannot pass it.

The scenario set build upon the first RiLSA example, but extends it by four traffic lights 200 m behind the evaluated traffic light. These traffic lights' green light durations are iterated between 8 s and 80 s in steps of 8 s for horizontal/vertical streams, respectively.



**Figure C.10: A comparison of the performance of a fixed-time and an actuated traffic light for the “RiLSA1Outflow” scenario set.**

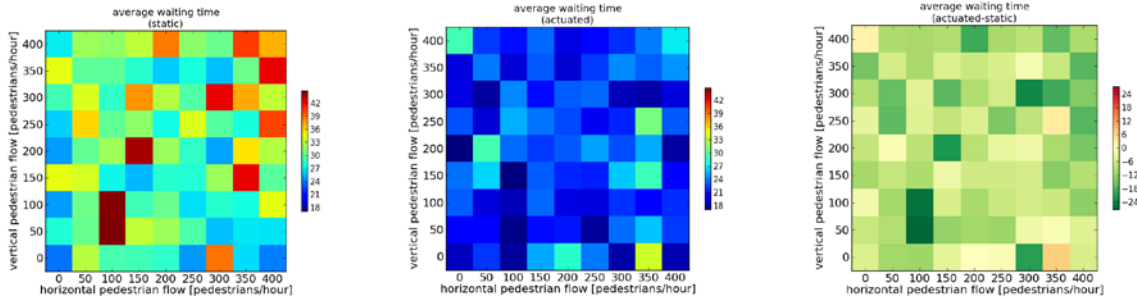
It can be clearly seen that traffic lights at the outgoing lanes influence waiting times to a high degree. The actuated traffic light performs worse than the fixed-time one, probably because the higher throughput of the major intersections yields in larger waiting times at the subsequent outflow reducing ones. This is in-line with the purpose of this scenario set: to determine whether jams at outgoing lanes are regarded. This is not the case for actuated traffic lights.

Notes about the scenario set:

- Some approaches are known to micro-control drivers' route choice by changing traffic light times – it is assumed that drivers will change the direction to continue the route depending on the green times ahead. Such a behaviour adaptation is not covered by this scenario set; it should be stated that no empirical analyses are known.
- The scenario targets on evaluating a traffic light that explicitly takes into account whether vehicles may leave the intersection. Such an algorithm is under development in COLOMBO, but has not been used for evaluations, yet.

### H: Pedestrian Flow (PedFlowIteration)

Most of the reports presented in [COLOMBO D1.1] use a simplified simulation model that does not include pedestrians. The “PedFlowIteration” scenario uses the basic RiLSA example 1 settings and adds pedestrian streams that interrupt the traffic by blocking the outflow lanes. The amount of pedestrians is iterated from 0 pedestrians/hour to 500 pedestrians/hour in steps of 100 pedestrians/hour for all directions.



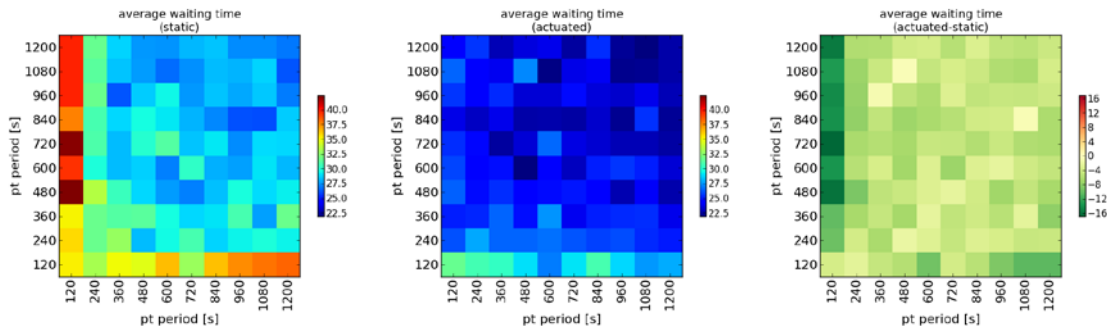
**Figure C.11: A comparison of the performance of a fixed-cycle and an actuated traffic light for the “PedFlowIteration” scenario set.**

Notes about the scenario set:

- As discussed, all specific characteristics besides the traffic demand, are currently covered by only one scenario. This is as well the case for taking pedestrians into account. As within COLOMBO’s Task 2.5 the developed traffic lights shall be extended by taking pedestrians into account, further scenario sets that address this topic will get necessary.
- The pedestrian model itself must be verified. As pedestrians have been included in the SUMO simulation recently, only few tests and evaluations have been performed, and a mature state may not yet been reached. Other pedestrian dynamic models are currently being included; the scenario set should be run with all of them.

### I: Public Transport (RiLSA1PTIteration)

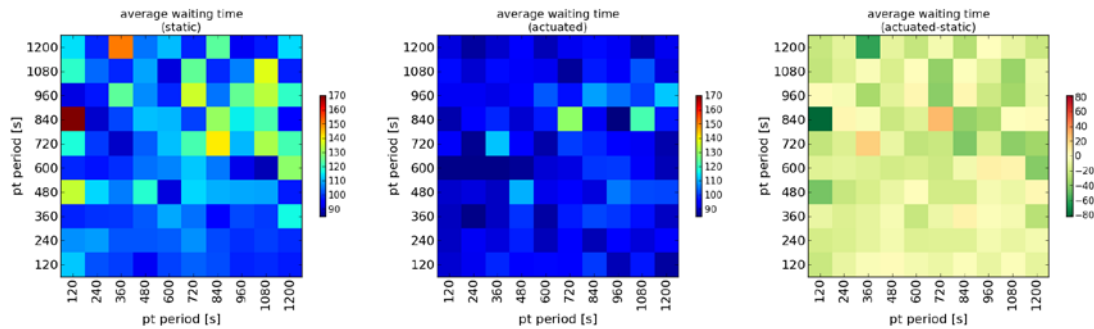
This scenario set targets on evaluating how a traffic light system prioritizes public transport. Again, the resampled first RiLSA example is used. In addition, public transport (busses) is injected. The injection period of public transport is iterated starting at 100 s to 1200 s in steps of 100 s. This is done for horizontal and vertical streams, respectively.



**Figure C.12: A comparison of the performance of a fixed-cycle and an actuated traffic light for the “RiLSA1PTIteration” scenario set.**

Figure C.12 shows that the additional busses influence the overall performance, especially if the injection happens often. Clear effects can be seen at an injection period of 120 s (every two minutes), but a gradient is still recognizable up to periods of 360 s length (six minutes). The remaining runs show a stochastic behaviour that cannot be explained as the performed twelve

iterations should straighten it. To investigate the effects of periods on bus schedules, the stored information about average travel time of vehicles of the type “bus” that are included within the runs’ database is used. It is shown in Figure C.13.



**Figure C.13: A comparison of the travel durations of busses for a fixed-cycle (left) and an actuated (right) traffic light.**

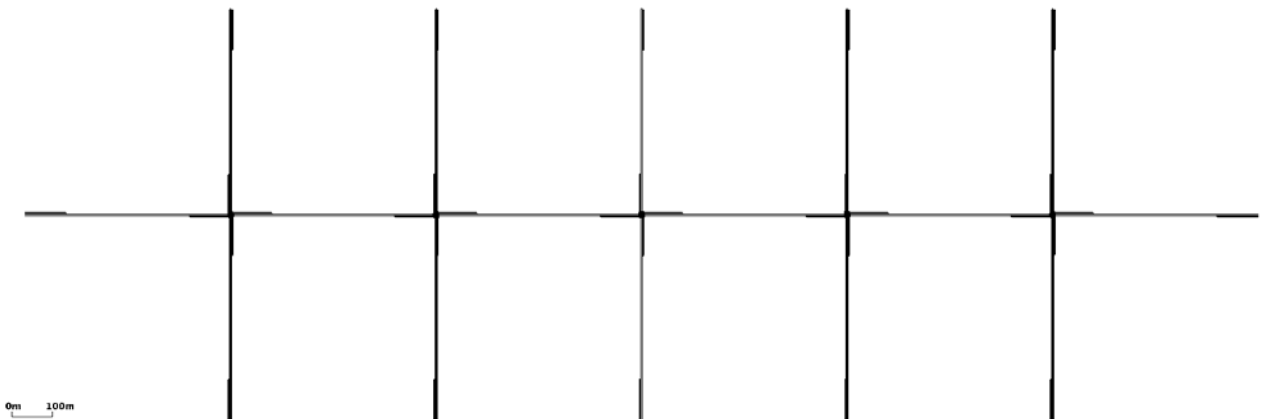
Even though the actuated traffic light compensates busses better than the one with a fixed cycle, the stochastic behaviour at lower public transport frequencies (higher periods) indicate that no explicit prioritisation takes place. If it would, higher durations may be expected at higher frequencies (because the traffic light cannot compensate more public transport) and the durations should form rather a plateau on lower frequencies. The observed behaviour is assumed to be correct, because none of the shown traffic light algorithms reacts on incoming public traffic in any means. A final evaluation of this scenario set is thereby only possible if an algorithm that performs public transport prioritisation is available – what is planned for later steps done in the COLOMBO project.

Notes about the scenario set:

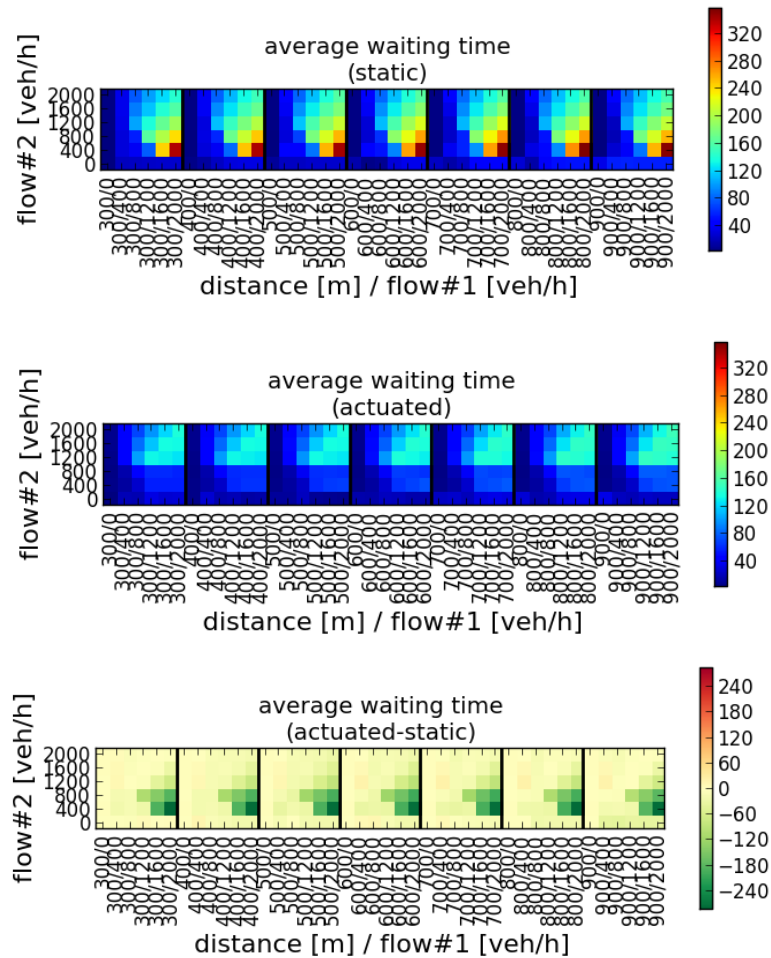
- One may think of testing the interaction between traffic lights and busses that start at a near-by bus stop. This could be easily obtained by adding a bus stop to the given scenario.

#### *J: Corridor Static Flow and Distance Changes (CorrFlowsDistancesA)*

The “CorrFlowsA” scenario set replicates demand changes as performed within the “iterateFlowsA”, including the adaptation of the traffic light, but using a road network that consists of five intersections in a row. All intersections are set up as the initial one depicted in Figure C.1. The distance between the intersections’ centres varied between 300 m and 900 m in steps of 100 m. The complete road network is shown in Figure C.14. All vertical flows have all the same traffic amount. This scenario set evaluates whether the traffic light algorithm coordinates the traffic lights.



**Figure C.14: The road network used for the “CorrFlowsA” scenario set. This image shows the network with a distance of 500 m between the intersections. Every intersection looks like the one depicted in Figure C.1.**



**Figure C.15: A comparison of the performance of a fixed-cycle and an actuated traffic light for the “CorrFlowsA” scenario set.**

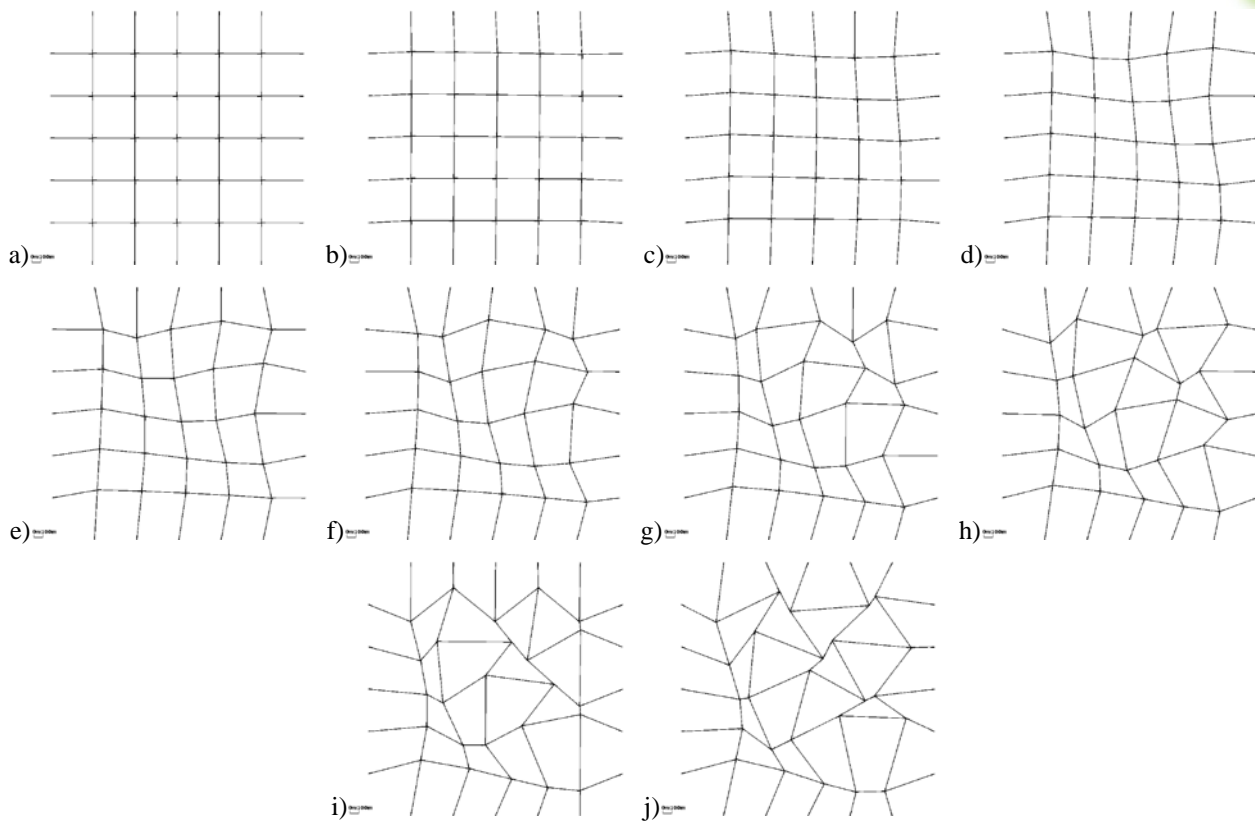
In both cases, the amount of vehicles is determining the performance in major. The distance seems to play a minor role. This may be side-effect of using an actuated algorithm that does not try to preserve same cycle times for coordination. To determine whether this scenario set is really valid requires further investigations.

Notes about the scenario set:

- A similar scenario set with no TLS adaptation is not needed as the evaluated TLC’s capabilities to adapt to a flow combination is already covered by “iterateFlowsNA”.
- The scenario does not cover any specific a-priori assumptions, such as predictions of a public transport vehicle’s arrival at subsequent intersections, e.g. that may be a matter of investigations. It does but cover implicit, “emerging” synchronisation as may happen when using traffic lights developed in COLOMBO’s WP2.

#### ***K: Network Static Flow and Distance Changes (NetFlowsDistancesA)***

A further attempt to determine how well a traffic light algorithm synchronizes traffic lights extends the scenario size to a network. To validate the synchronisation abilities of the investigated traffic light for different network layouts, an additional iteration is used within which the positions of the network intersections are modified. The resulting networks are as shown in Figure C.16.



**Figure C.16: Scenario networks.**

Again, the major difference between the compared algorithms is actuated traffic lights' possibility to react on traffic flow differences, as already shown in the "iterateFlowsA" scenario set. Changes in the network shape as done along the "offset"-axis do not show a major influence. Whether the scenario set is useful can thereby be only determined when investigating an algorithm that directly targets on synchronisation.

### *L: Real World Scenarios*

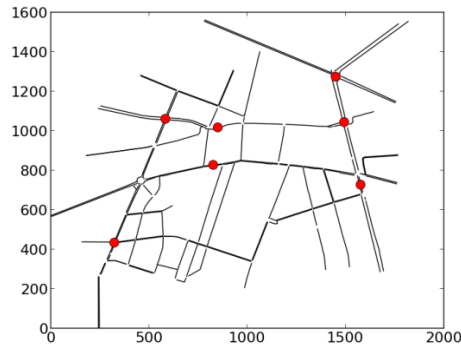
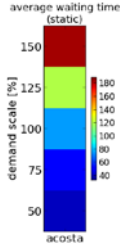
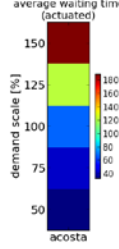
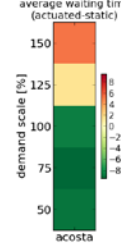
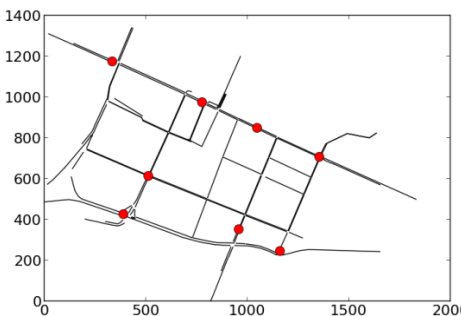
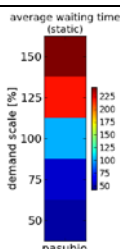
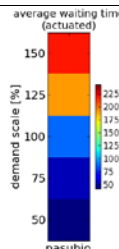
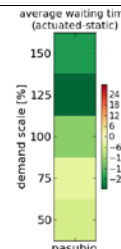
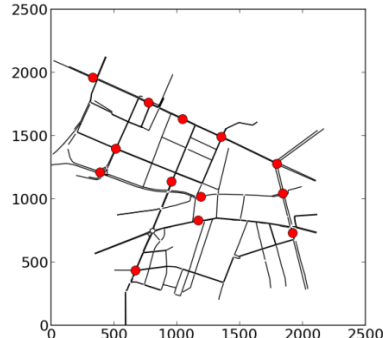
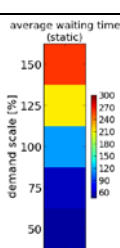
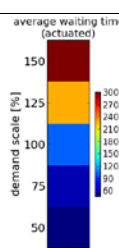
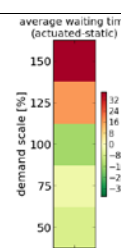
The aforementioned scenario sets pose the evaluated TLC against a large variety of problems to solve. Still, all these settings are artificial and target to determine the control's performance under one or a small subset of possible influences. Real-world settings but may include more influences and different inter-dependencies between them than what has been tested so far. Therefore, to obtain a complete view at a traffic light control's performance, additional representations of real-world scenarios are included in CTAES. Most of them were already presented in D1.1 and are herein only briefly described.

For a named real-world scenario, the scenario set iterates over demand scales starting at 50 % of the originally given demand, to 150 % in steps of 25 %, a method is known in literature. Scenarios released by COLOMBO are included within the scenario sets and their results regarding the performance of fixed-time and actuated traffic lights is given in Table C.1<sup>14</sup>.

<sup>14</sup> Only one simulation run per scale was performed for the "Pasubio" scenario



Table C.1: Behaviour of released real-world scenarios.

Scenario	Network and TLS	Performance (avg. waiting time)
ACosta		  
Pasubio		  
joined		  

It is interesting to note that the actuated traffic light behaves worse than the fixed-cycle one within the “A. Costa” and the “joined” scenarios at high flows. The definite reasons have yet to be investigated.

Notes about the scenario set:

- Usually, such the approach for scaling a demand is applied using non-peak hour traffic; this should be revalidated for the given scenarios.